UNIVERSITY - NATIONAL OCEANOGRAPHIC LABORATORY SYSTEM

RESEARCH VESSEL TECHNICAL ENHANCEMENT COMMITTEE

MEETING MINUTES

September 19-21, 1993

Martin Johnson House (Bldg T-29)
Scripps Institution of Oceanography
LaJolla, CA

RESEARCH VESSEL TECHNICAL ENHANCEMENT COMMITTEE
MARTIN JOHNSON HOUSE (BLDG T-29)
SCRIPPS INSTITUTION OF OCEANOGRAPHY
LA JOLLA, CA
SEPTEMBER 19 - 21, 1993
MEETING MINUTES

SCRIPPS TOUR: The RVTEC convened on Sunday, September 19, 1993 at Scripps in La Jolla, CA. David Wirth provided a tour of the campus.

Appendices

I. Meeting Agenda
II. RVTEC Meeting Attendance List
III. RVTEC By-Laws
IV. CTD Flow Charts and Data
V. CTD References
VI. Technical Personnel Listing
VII. Technician Exchange Policy Statement
VIII. Data Format Information
IX. Data Distribution Information
X. ADCP Issues
XI. ADCP Data Examples
XII. SeaNet Shipboard LAN Guide

INTRODUCTION: The meeting was called to order at 0830 September 20, 1993 by Chair, Rich Findley. Rich reviewed the agenda which is included as Appendix I. The meeting participant list is included as Appendix II.

ELECTIONS: The term of office for the first Vice Chair had been limited to one year so that the elections of the Chair and Vice Chair would occur in alternate years. The incumbent Vice Chair, Doug Biggs, declined to accept renomination and Mark Willis and Tim Pfeiffer were nominated. Tim Pfeiffer was elected to serve as Vice Chair. Doug Biggs was thanked for his service as Vice Chair.

ADOPTION OF ANNEX: The RVTEC by laws (see Appendix III) were approved for adoption as Annex V to the UNOLS Charter. They will be forwarded to the UNOLS Council and Membership for endorsement.

CTD WORKSHOP: Bob Millard and Jim Swift then began the workshop on CTD calibration and data processing. Documentation and record keeping are critically important for maintaining the quality of CTD data.

A salinity offset of 0.001 psu can be caused by an offset of:

+/- 0.001 degrees c in temperature
+/- 0.001 ms/cm conductivity
+/- 2.5 db in pressure

An offset of 0.01 ml/l in oxygen can be caused by an offset of:

+/- 0.25 degrees c in temperature
+/- 0.25 psu in salinity
+/- 50 db in pressure

The WOCE operations manual includes chapters on processing CTD data and will be available shortly. The RVTEC mailing list will be added to the WOCE mailing list and members of RVTEC will receive copies of the Operations Manual.

Bob discussed the software he uses for processing CTD data; flow charts are included in Appendix IV and Bob has agreed to make this software available.

There was some discussion of submission of data to NODC and while some of the support groups offer this as a service, it is actually the responsibility of the scientists to make the submission.

There are different levels of use of CTD data by scientists from different disciplines. The following points emerged during the ensuing discussion:

1. Data is being collected which is not ultimately archived.
2. Calibration procedures require resources such as technician time, wire time, and standard sea water which the Chief Scientist may wish to allocate differently.
3. Some mechanism is needed for documenting the calibration level or standard to which CTD data was collected.
4. Calibration to full WOCE performance specifications is probably not practical in shallow coastal waters.

Bob provided a bibliography which is included as Appendix V.

The meeting was adjourned for lunch at 1200 and reconvened at 1300.

CATALOGING MAILING LISTS/DATABASE SERVICES: Tom Wilson presented the report of the technician and equipment database subcommittee. He distributed equipment lists and technician biographies from several institutions. The format in which the information was collected is included as Appendix VI. Discussion of the most appropriate form for making this information available followed and the consensus was that a flat ASCII file which could be searched by a variety of tools was the most appropriate. Some degree of standard nomenclature would be imposed during the final editing process. It was also agreed that the information should be available through Omnet or the Internet. It was also recommended that a catalog be published and distributed through the UNOLS Office. John Freitag moved that Tom Wilson be asked to continue the work of the Technician Database subcommittee. The motion was seconded by Steve Rabalais and passed unanimously.

TECHNICIAN TRAINING/TECHNICIAN EXCHANGE: Steve Rabalais presented the report on the progress of the Technician Exchange and Training subcommittee. Steve reported that a substantial number of exchanges had taken place last year, primarily driven by economics. John Freitag reported that with ENDEAVOR undergoing a mid life refit he had supplied technicians to Bermuda, the R/V CAPE HATTERAS, and L-DEO. After some discussion of the administrative details of compensation and working conditions a consensus was reached that:

1. The administrative details can be made transparent to the borrowing institutions and can be worked out on an ad hoc basis between the two institutions concerned and Lisa Rom's office.
2. Our first year's experience with this effort indicates that sharing technicians between institutions is possible to do, is of benefit to both institutions, and the program can be continued as is with no further formal discussion required.

David Asselin of Antarctic Support Associates said that he needed technicians for work in the Antarctic for all of next year. Further discussion revealed that the administrative details were more complex than in exchanges between UNOLS institutions. Lisa agreed to examine this matter at NSF.

It was recommended that speakers, perhaps industry representatives, be invited to the next meeting to provide a training session. A motion was made by John Freitag to revise the existing charge to the subcommittee on Technician Training/Exchange to concentrate on educational issues and establishment of training sessions at future RVTEC meetings. The motion was seconded by Frank White and passed unanimously.

Steve moved that the draft policy statement on technician training, which is included as Appendix VII, be accepted. The motion was seconded by Doug Biggs and passed unanimously.

NEW INSTRUMENTATION: The subcommittee for new instrumentation was dissolved.

DATA STANDARDS: Barrie Walden, WHOI, presented a report which contrasted simple ASCII files to more complex data files. ASCII files are simple, access to the data is immediate, frequently requiring software no more complex than a text editor or spreadsheet. However, documentation is lacking or minimal. As more complete documentation is included in the files, then the simplicity of access to the files is lost and if the documentation is in separate files then it eventually gets separated from the data.

The alternative is a more complex data storage format which produces true self documenting files. These are typically binary files and are organized to contain both the data itself and all the attributes of that data needed to interpret it. Specific programs are required to both read and write the files. NetCDF was proposed as one possible candidate for use by the RVTEC community. Rich Findley provided a set of handouts on various data formats; these are included as Appendix VIII. Since most of the group had no direct experience with NetCDF or other data standards the ensuing discussion was oriented more toward the pros and cons of employing any data standard, rather than the specifics of NetCDF versus some other format. The following points emerged from the discussion:

- The intent of this effort is to develop tools which will allow us to assist each other with the issues of the management of data quality, documentation, and calibration. It is not our intent to enforce a data standard which would be imposed on the scientific users of the data.

- Many of us currently provide data to users in flat ASCII files. Tools for NetCDF exist which will allow us to extract data from the NetCDF files and create flat files similar to the ones we now provide.

- If the oceanographic community does in the future move toward adaptation of a truly standard data standard which is different from whatever RVTEC adopts now then it will be a much simpler task to convert from, for instance, NetCDF to the new standard than it would be to convert from our present mixture of formats. It can be referred to as the UNOLS Data Interchange Format (UDIF).

Rich Findley then began a discussion of the use of CD-R technology by summarizing the points from his "The problem from where I sit..." posting to UNOLS.RVTEC, which is included as Appendix IX. There was complete agreement that CD-R drives have great potential for solving the hardware side of the data dissemination and archiving problems we face. The one drawback appears to be that the CD-R drives appear to be too sensitive to motion to be able to write disks at sea. The CD-ROM readers, on the other hand, are all ready in routine use on several ships and function well at sea.

Mark Willis agreed to lead a subcommittee to further evaluate various data standards for use by RVTEC. Dave Nelson will attend a workshop at URI on oceanographic data standards and Marty Mulhern agreed to look into NOAA's progress in this field.

The first meeting day was adjourned at approximately 5:00 p.m.

The meeting was reconvened by Chair Rich Findley at 0830 Tuesday September 21, 1993.

ADCP WORKSHOP: Teri Chereskin and Eric Firing led a workshop on ADCP's. During her introductory remarks Teri emphasized that documentation of calibration and maintenance procedures is important. Since ADCP's require little attention once they are properly installed, what is done tends to not be documented. A check list of points to include in the documentation is included as Appendix X. Teri also provided *WHP Underway Measurements Guidelines* for distribution by the UNOLS Office. Eric discussed the results of his work with the Ashtec 3DF GPS heading sensor. Plots of the gyro error compared to the Ashtec are included as Appendix XI. Eric emphasized correcting the gyro compass with a GPS heading sensor is essential for quality ADCP work. We discussed the problem of processing and archiving ADCP data on cruises where it is not of primary interest to the scientific party. Eric pointed out that there is an "XBT Effect" in that as the amount of ADCP data available increases it becomes more valuable. Eric discussed the possibility of establishing one or more regional centers to process and archive ADCP data.

The meeting was adjourned at 1200 for lunch and reconvened at 1300.

SEANET PRESENTATION: Ellen Kappel of JOI; Rex Buddenberg, JOI; and Andy Maffei, WHOI; led a discussion of SeaNet. Ellen began with the background of the SeaNet program which is intended to ultimately bring a proposal to NSF concerning the connection of all UNOLS vessels to the Internet. The goal is to establish high-speed data communication transfers; allowing real-time data distribution. Rex discussed the SeaNet Shipboard LAN Guide, which is included as Appendix XII, and the extension of the terrestrial wide area networking currently supported by NSF in the Internet to radio, satellite based communications. Rex also discussed the advancements in communications facilities which will be available in the near future. Some low earth orbit (LEO) satellites are being deployed now and will provide store and forward e-mail type services shortly. Networks of large LEO's will be deployed six years from now which will provide the bandwidth and connectivity to allow ships to connect to the Internet as simply as a system currently connects on a campus. In the discussion that followed there was a consensus that reliable e-mail services would fill many of our current needs.

NEW BUSINESS: At the conclusion of the data communication and SeaNet discussion, Rich opened the meeting to new business.

Doug Biggs mentioned that there would be a workshop in December sponsored by IOC on Marine Chemical Measurements. Doug will report back on this.

Andy Maffei agreed to investigate setting up an FTP server for our use at WHOI.

Annette reminded us that UNOLS would have a booth at the AGU meeting.

Steve Rabalais agreed to continue as RVOC liaison.

Don Moller reported that the funding for the cable pool wire purchases, both 3X19 and EM cable, is progressing.

We agreed to include a half day workshop on wire, cable, and terminations at the next meeting.

The meeting was adjourned at 1600.

TENTATIVE AGENDA

RESEARCH VESSEL TECHNICAL ENHANCEMENT COMMITTEE
SEPTEMBER 19, 20, 21, 1993
MARTIN JOHNSON HOUSE (Bldg. T-29)
SCRIPPS INSTITUTION OF OCEANOGRAPHY
LaJOLLA, CA

Sunday

2:00	Meet at Martin Johnson House (Bldg. T-29) on SIO Campus. Group will then depart for tour of Nimitz Marine Facility.

Monday

8:30	Meeting Called to Order - Introductory Remarks by Chair
8:45	Participant Introductions
9:00	Election of Vice Chair
9:15	Adoption of Annex to UNOLS Charter
9:30	CTD Guest Speakers; Bob Millard, Jim Swift followed by CTD Round Table Discussions
10:15	***Break***
10:30	CTD Round Table Discussions (cont.)
12:00	***Lunch***
1:00	Subcommittee Reports and Working Discussions (40 minutes each) Catalog Mailing Lists/Database Services; Tom Wilson Technician Training/Technician Exchange; Steve Rabalais
2:30	***Break***
2:45	New Instrumentation; (looking for a chair)
3:00	Data Standards followed by Round Table Discussion Data Standards
5:00	***Adjournment***

(Continued on back)

Tuesday

8:30	Meeting Called to Order ADCP Guest Speakers; Eric Firing and Teri Chereskin followed by ADCP Round Table Discussions
10:00	*Break*
10:15	ADCP Round Table Discussions (cont.)
12:00	*Lunch*
1:00	Presentation and Discussion of SeaNet; Ellen Kappel
3:00	*Break*
3:15	Development of Goals and Objectives
3:30	Updating of Action Plans
3:45	Confirmation of Subcommittees
4:00	Scheduling of Next Meeting
4:15	New Business
5:00	*Adjournment*

RVTEC ATTENDANCE LIST

NAME	AFFILIATION
David Asselin	Antarctic Support Association
Doug Biggs	Texas A&M
Barry Bjork	Bermuda Biological Station
Rex Buddenberg	SEANET
Teresa Chereskin	SIO
Annette DeSilva	U Rhode Island
James Donovan	LUMCON
Rich Findley	U Miami, RSMAS
Eric Firing	U Hawaii
John Freitag	U Rhode Island
Andy Heard	MLML
Ellen Kappel	JOI
Frank Kelly	Texas A&M
Andrew Maffei	WHOI
Robert Millard	WHOI
Don Moller	WHOI
Marty Mulhern	NOAA
David Nelson	U Rhode Island
Tim Pfeiffer	U Delaware
Steve Poulos	U Hawaii
Steve Rabalais	LUMCON
Mike Rawson	L-DEO
Lisa Rom	NSF
James Swift	SIO
Neal Thayer	USCG
Barrie Walden	WHOI
Bob Luke Wejinya	U Michigan
George White	U Washington
Marc Willis	OSU
Tom Wilson	SUNY, Stony Brook
David Wirth	SIO

BYLAWS OF THE RESEARCH VESSEL TECHNICAL ENHANCEMENT COMMITTEE

A. PURPOSE

1. The purpose of the Research Vessel Technical Enhancement Committee shall be to promote the scientific productivity of research programs that make use of research vessels and oceanographic facilities and to foster activities that will lead to enhanced technical support for sea-going scientific programs.

B. MEMBERSHIP

1. Membership in the RVTEC shall be extended to UNOLS member institutions.

2. Participation shall be open to technical and scientific personnel at UNOLS and non-UNOLS organizations.

C. REPRESENTATION

1. Each institutional UNOLS representative may designate a representative to RVTEC.

2. RVTEC will meet at least once per year.

3. The place and time of the next RVTEC annual meeting will be designated at the close of the previous RVTEC meeting.

4. Each member institution shall be notified of the next annual meeting by the Vice Chairperson of the Committee at least 90 days prior to the next annual meeting.

5. Each UNOLS member institution shall be entitled to one vote on matters at RVTEC meetings. However, matters may be submitted for vote by the Chairperson at other times. These matters will be voted on by mail or electronic mail, and votes will be collected for a period of two weeks.

6. A simple majority of the UNOLS operator institutions must be represented to establish a quorum.

D. OFFICERS

1. The Research Vessel Technical Enhancement Committee shall have a Chairperson and a Vice Chairperson. The Chairperson and Vice Chairperson will be elected by majority vote at the Annual Meeting and subject to confirmation by the UNOLS Chair. Their terms of office shall be two years. The Chairperson and Vice Chairperson shall be elected in alternate years.

2. The Chairperson shall represent the Committee in all matters stipulated in the purpose of these bylaws and in all matters deemed necessary in the interest of the Committee. The Chairperson shall implement the programs enumerated by the Committee and shall conduct the Annual Meeting and whatever special meetings are deemed necessary by the Chairperson or the members.

3. The Vice Chairperson, who shall function as Chair in the absence of the Chair, shall be responsible for recording the business of the Committee and for dissemination of information through a newsletter or other media as stipulated in these bylaws to all members of the Committee.

4. If the Chairperson or Vice Chairperson are unable to fulfill their duties of office, the Chairperson shall appoint a successor to act with authority until the succeeding Annual Meeting.

E. WORKING GROUPS AND PANELS

1. Upon the recommendation of the Chairperson, and with a majority vote of the Committee, various working groups and panels, as necessary to the work of the Committee, may be constituted. The duration of action of such working groups and panels shall be stipulated at the time of inception.

F. MEETINGS

1. A general meeting of the Committee shall be held at least once yearly. The Chairperson shall preside over this Annual Meeting. The business of this meeting shall encompass reports of any active working groups and panels, and discussions of project and actions of the Committee. Research Scientists and others from the marine community may also be included on the agenda. Workshops for projects of general concern are encouraged.

2. Passage of projects and actions shall be by vote, in accordance with the voting procedures set forth in Section C, REPRESENTATION, paragraph 5.

3. The various working groups and panels shall each meet at least once yearly.

G. FINANCES

1. UNOLS will provide limited funding for the Committee to include the following:

   a. Travel expenses for the Chairperson and Vice Chairperson for meetings once a year;

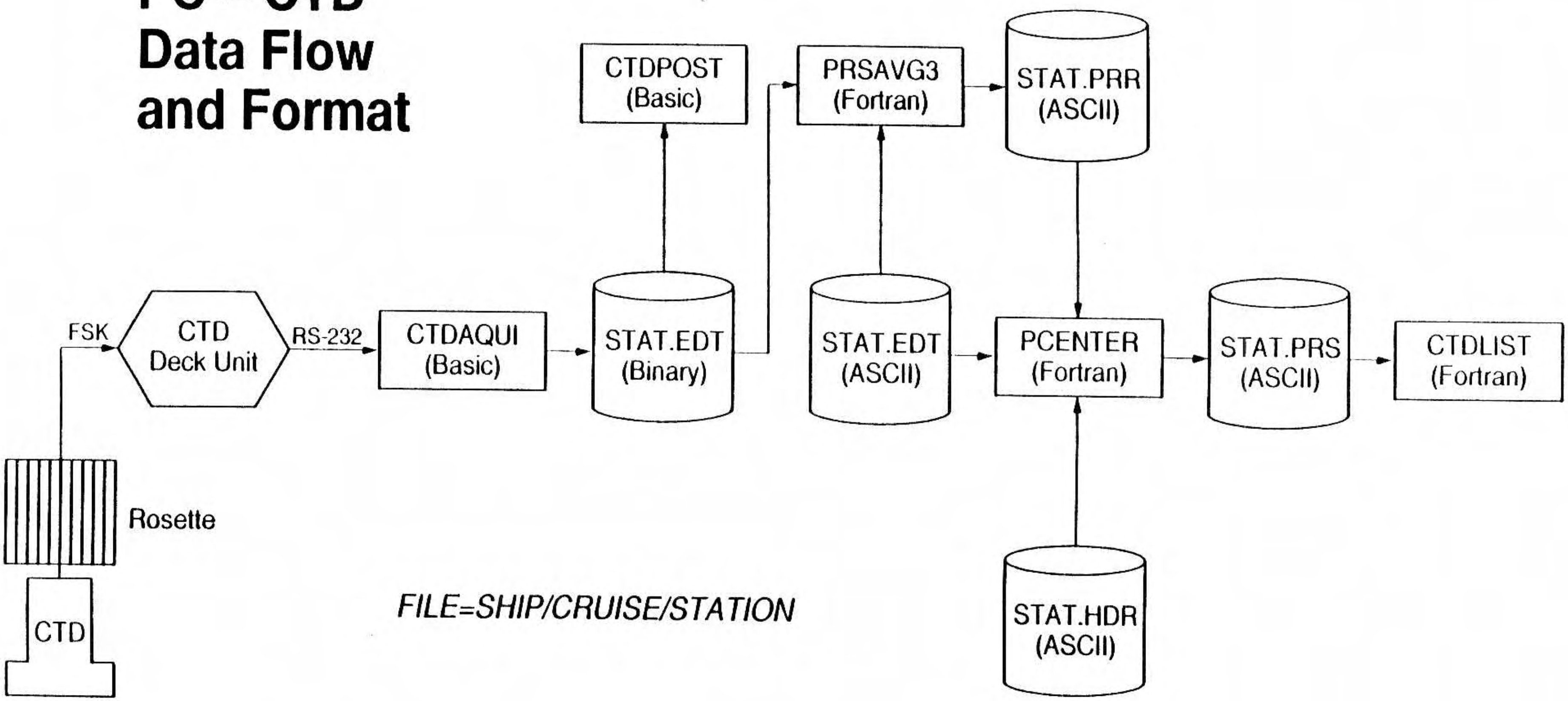
   b. Travel expenses for the Chairperson to attend UNOLS Meetings;

   c. Meeting facilities, when required;

   d. Travel and meeting expenses for panels, workshops, or the Annual Meeting when appropriate.

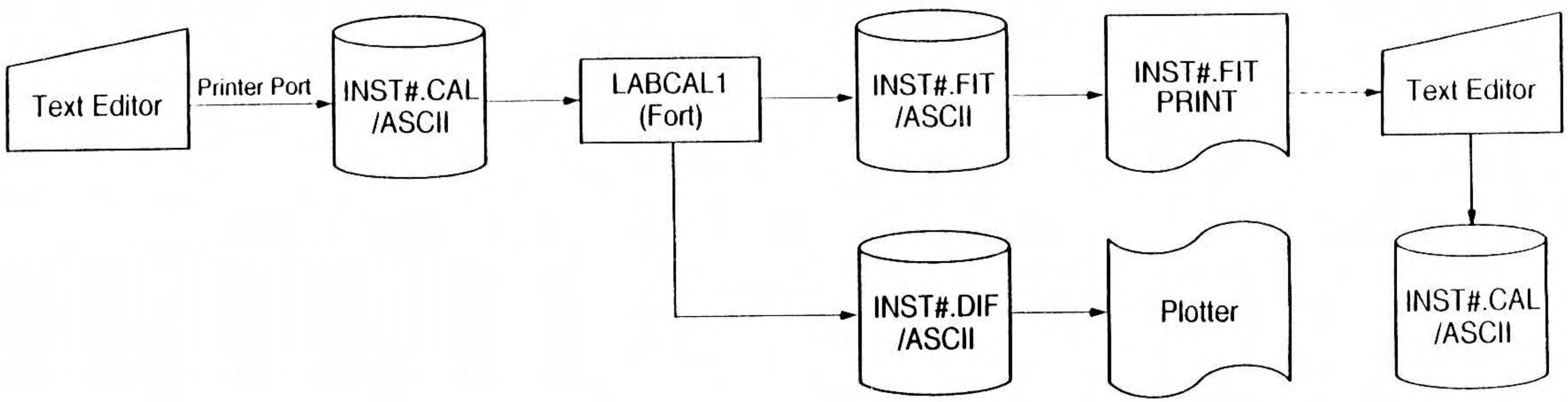
Approved and adopted on 18 October 1992 at the RVTEC Organizational Meeting in Washington, DC.

PC – CTD Data Flow and Format

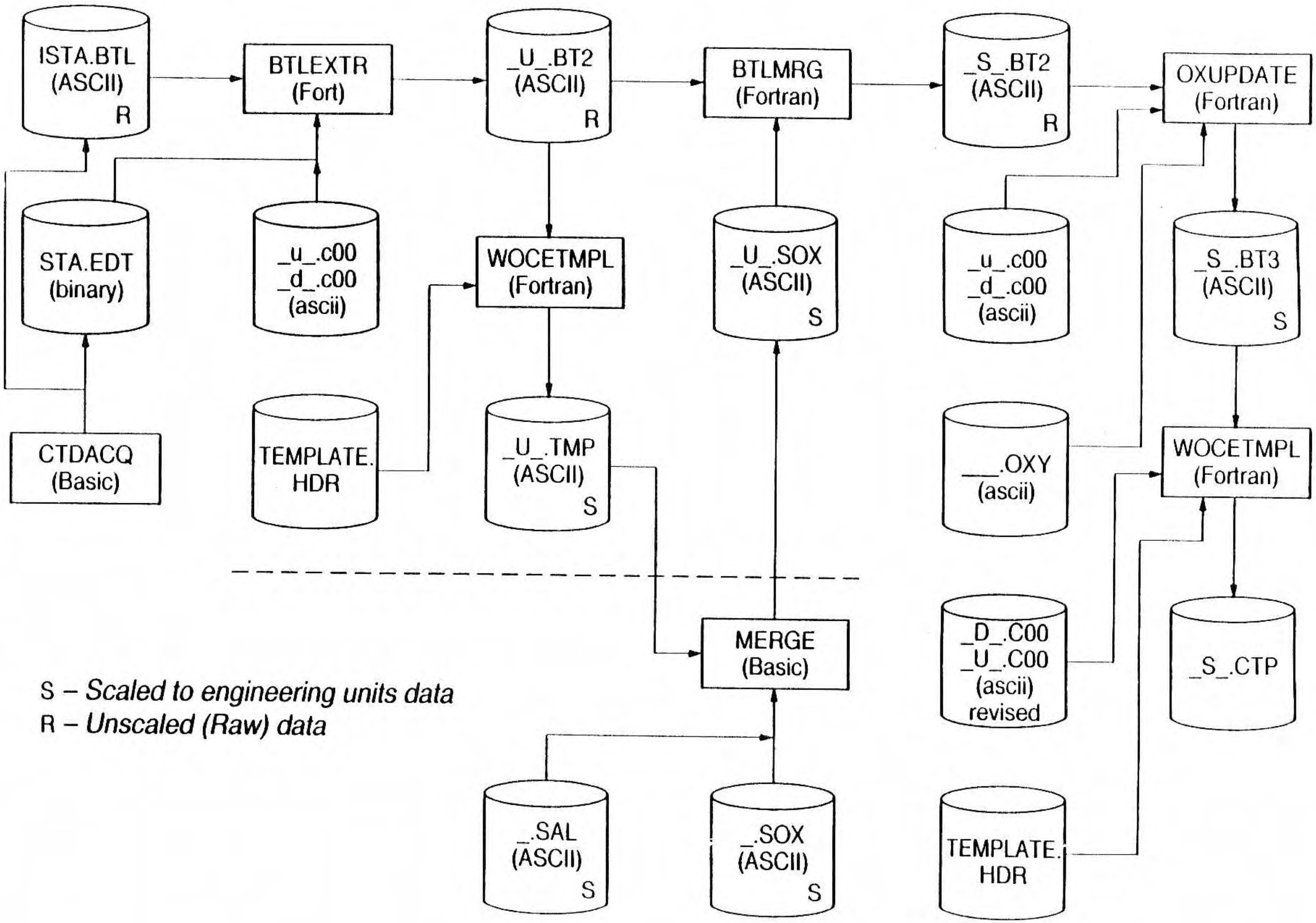


Laboratory Calibration Data

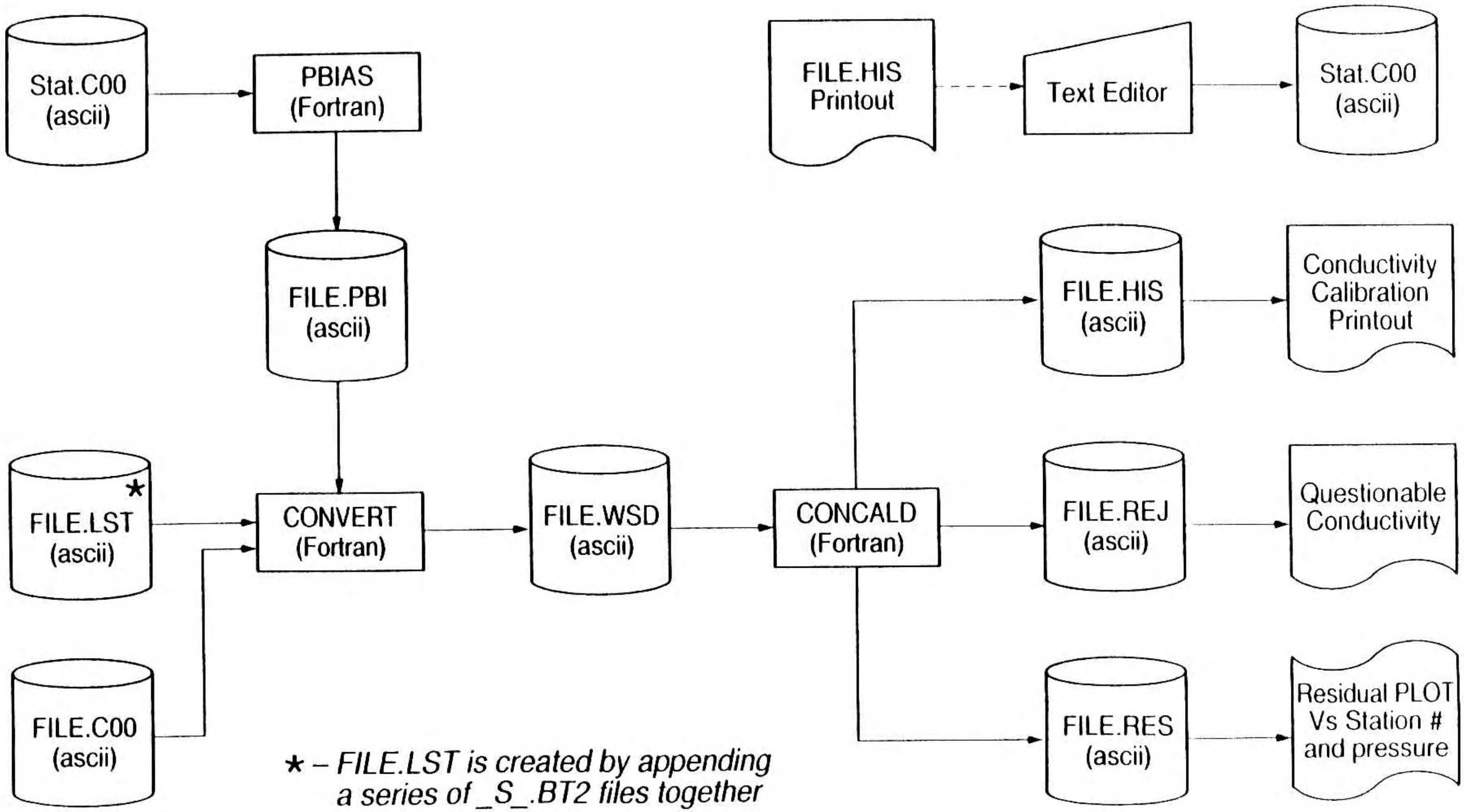
(CREATE INSTRUMENT CALIBRATION FILE)



Water Sample Merge

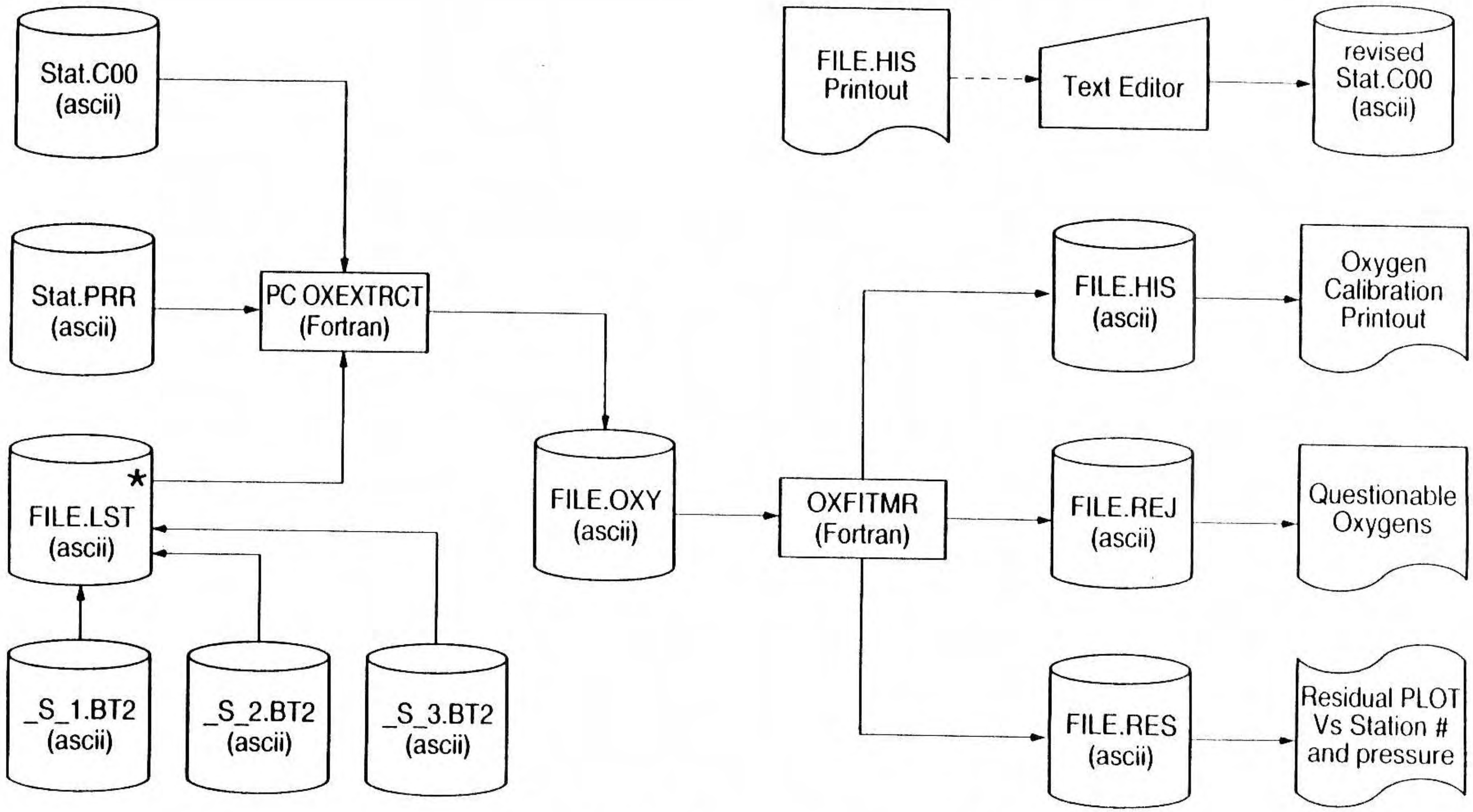


Calibration of CTD Conductivity (Salinity) to Water Samples



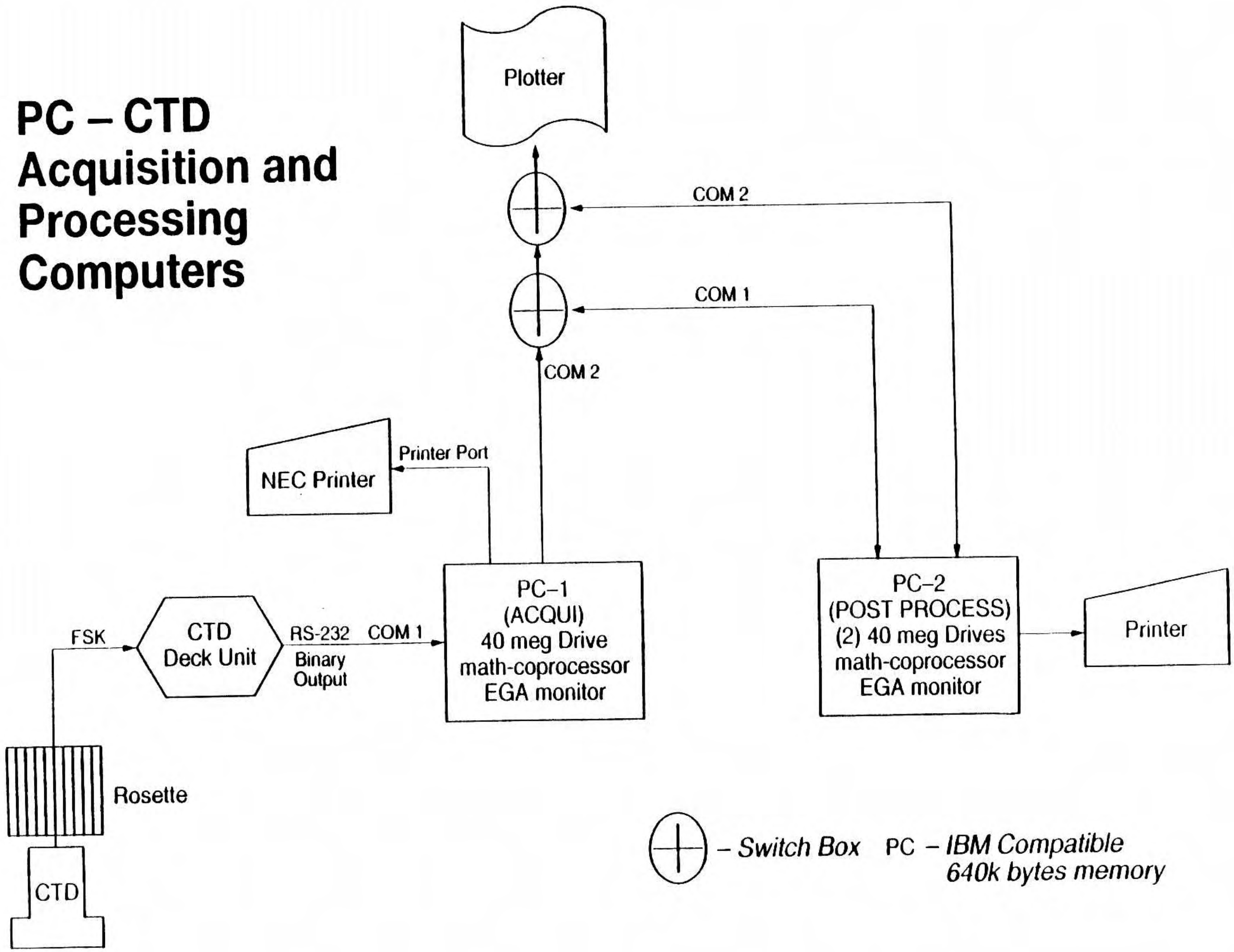
* – *FILE.LST is created by appending a series of _S_.BT2 files together with header*

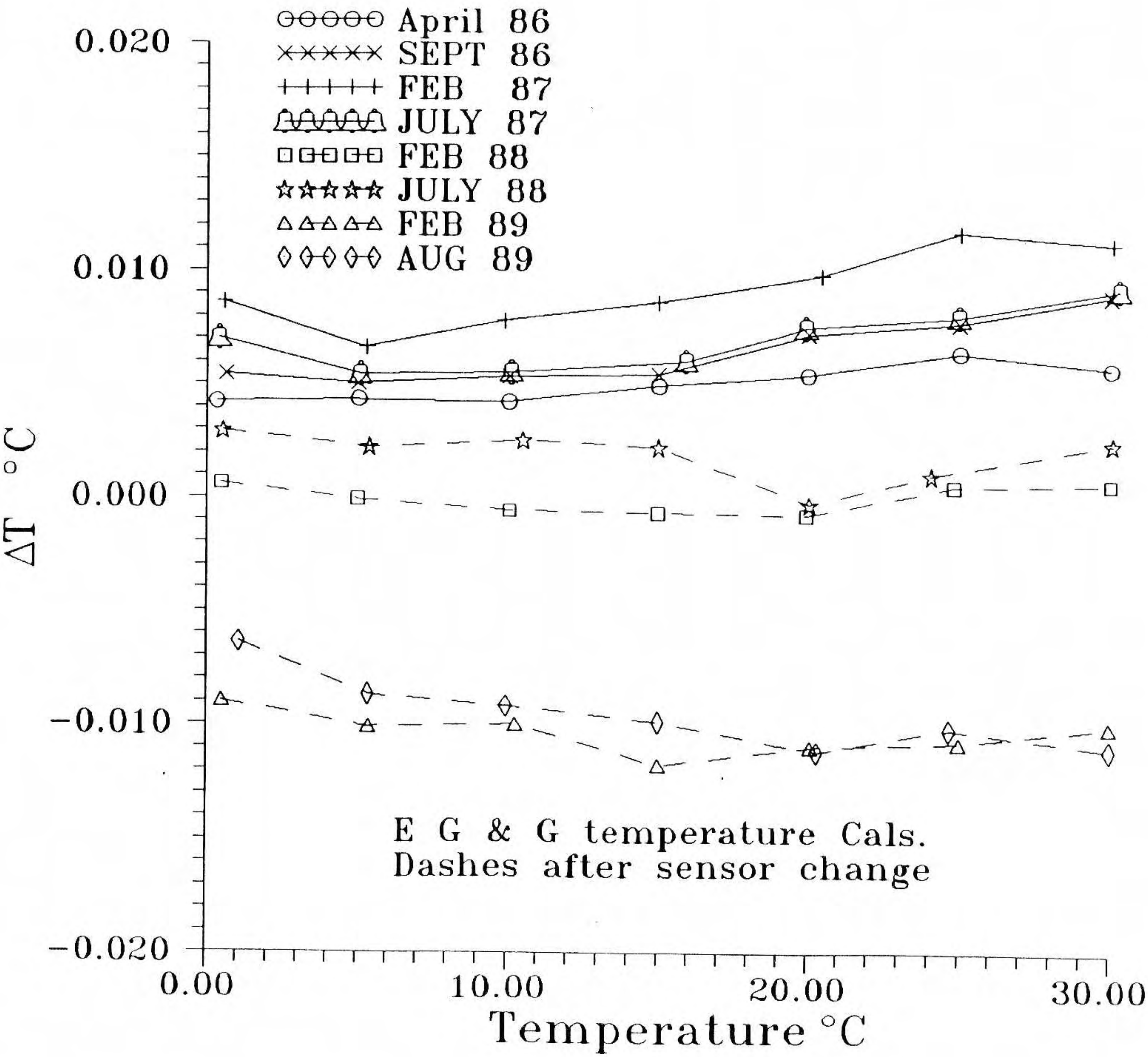
Calibration of CTD Oxygen with Water Samples



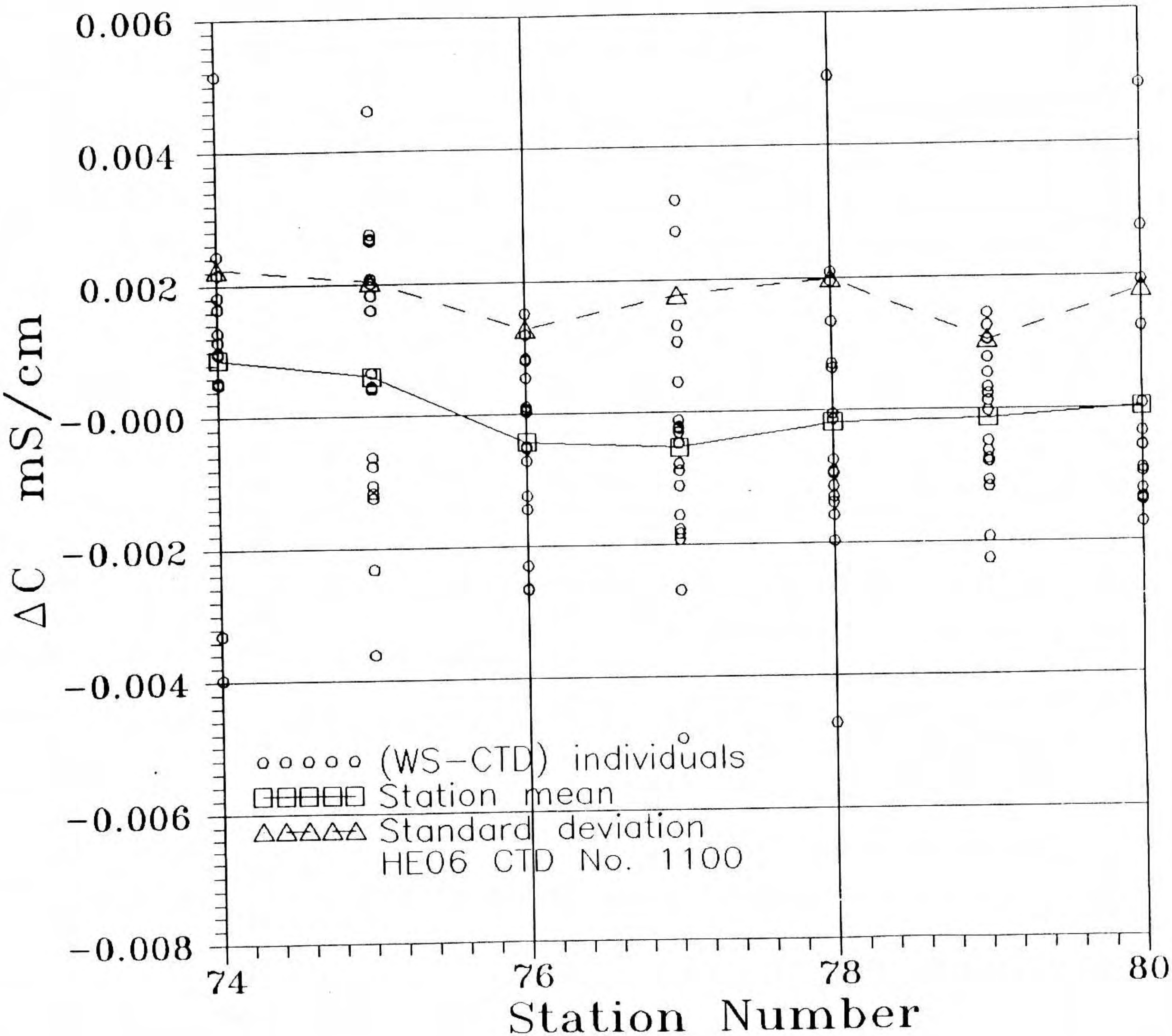
* – *FILE.LST is created by appending a series of _S_.BT2 files together with header*

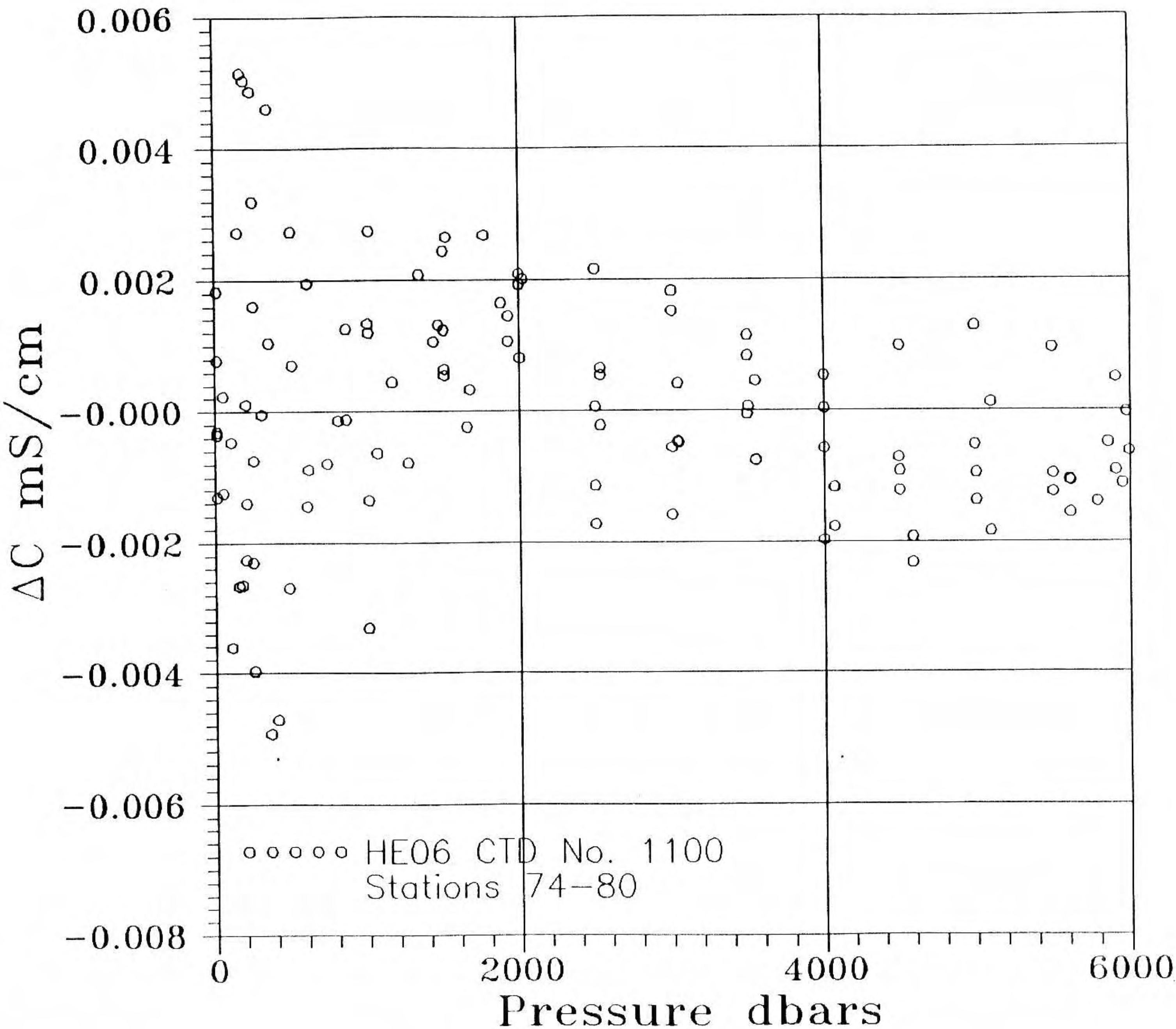
PC – CTD Acquisition and Processing Computers



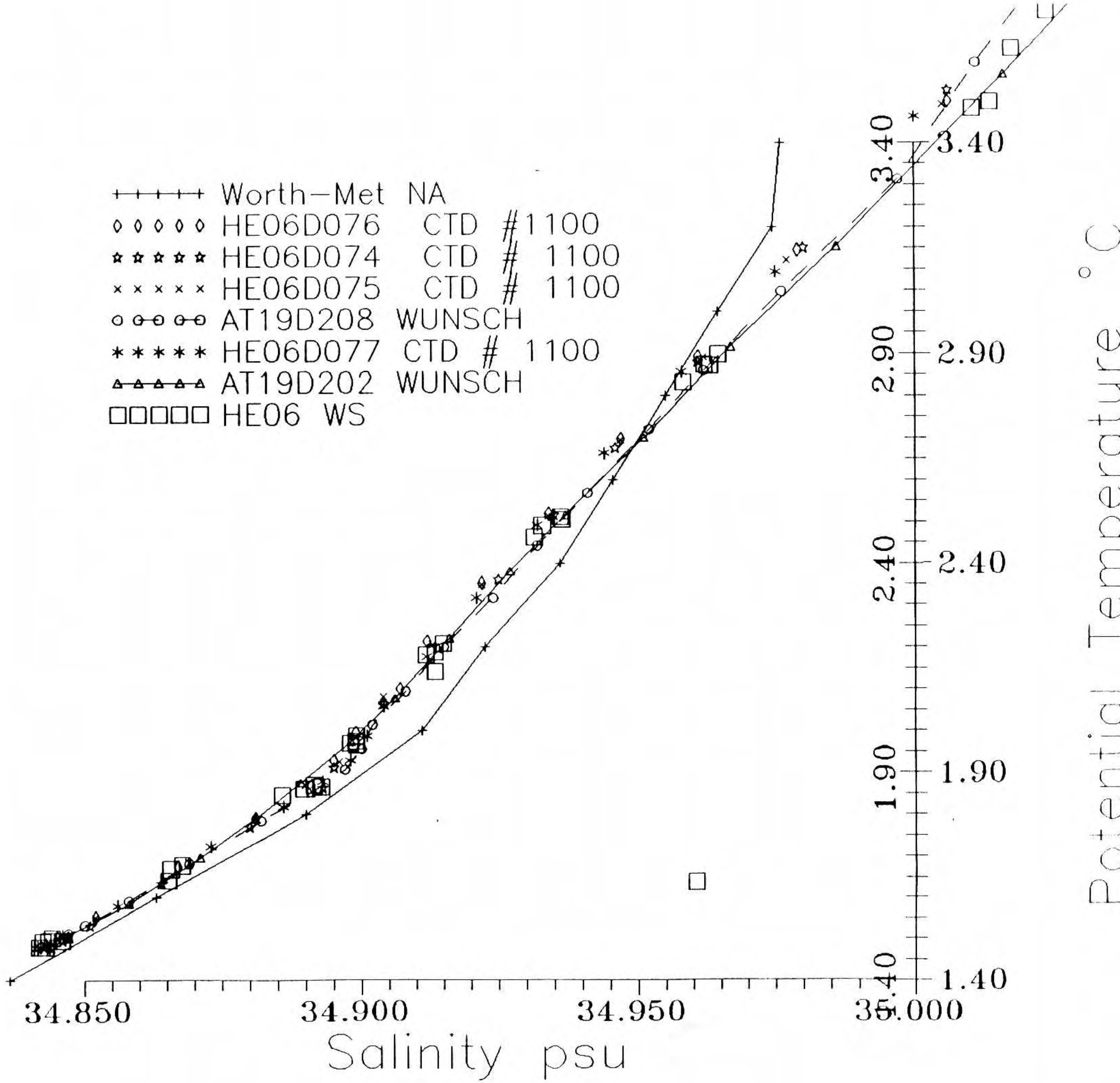


April 86
SEPT 86
FEB 87
JULY 87
FEB 88
JULY 88
FEB 89
AUG 89
0.020
0.010
0.000
-0.010
-0.020
ΔT °C
0.00
10.00
20.00
30.00
Temperature °C
E G & G temperature Cals.
Dashes after sensor change

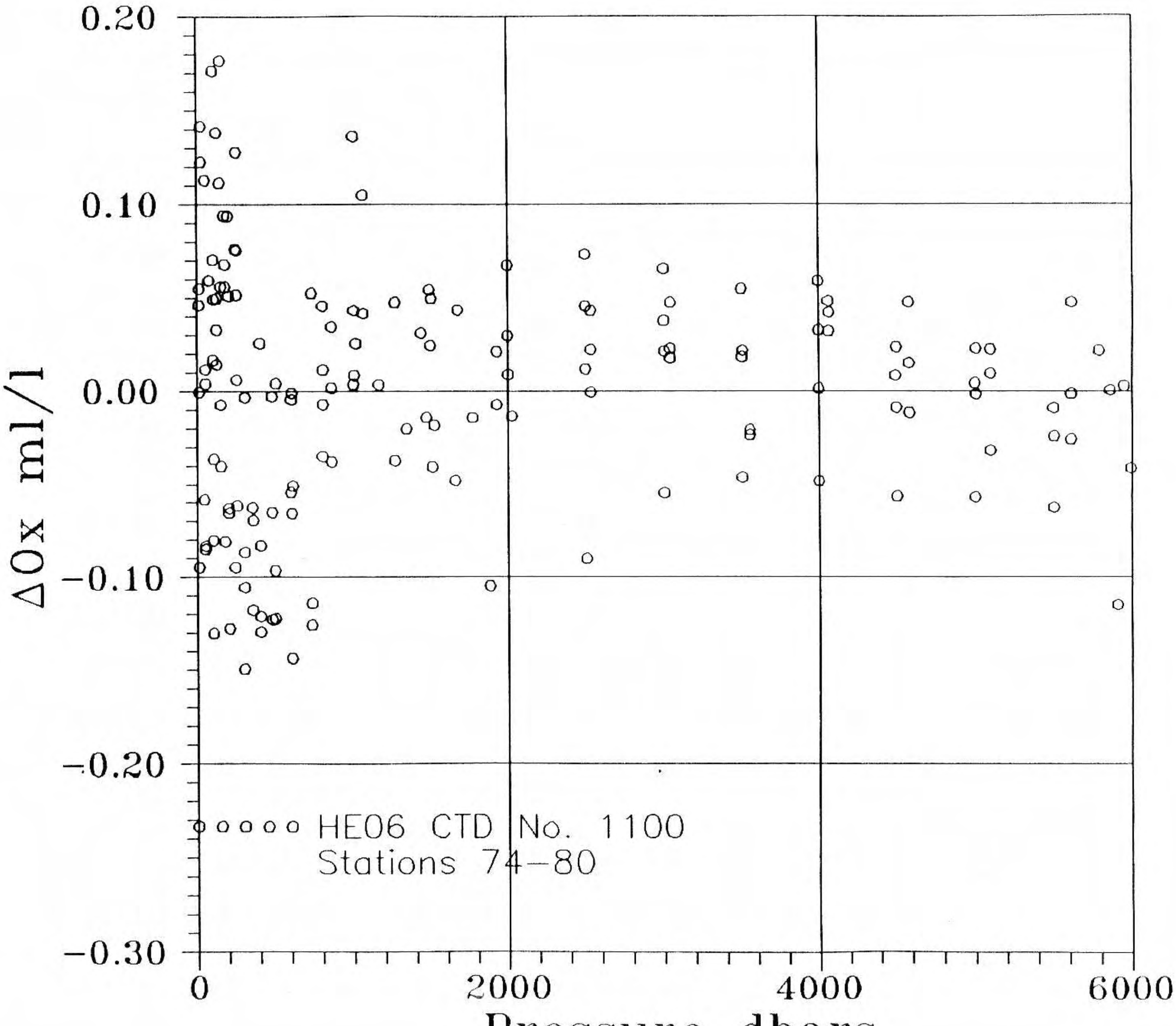

0.006
0.004
0.002
-0.000
-0.002
-0.004
-0.006
-0.008
ΔC mS/cm
74
76
78
80
Station Number
○○○○○ (WS-CTD) individuals
□□□□□ Station mean
△△△△△ Standard deviation
HE06 CTD No. 1100

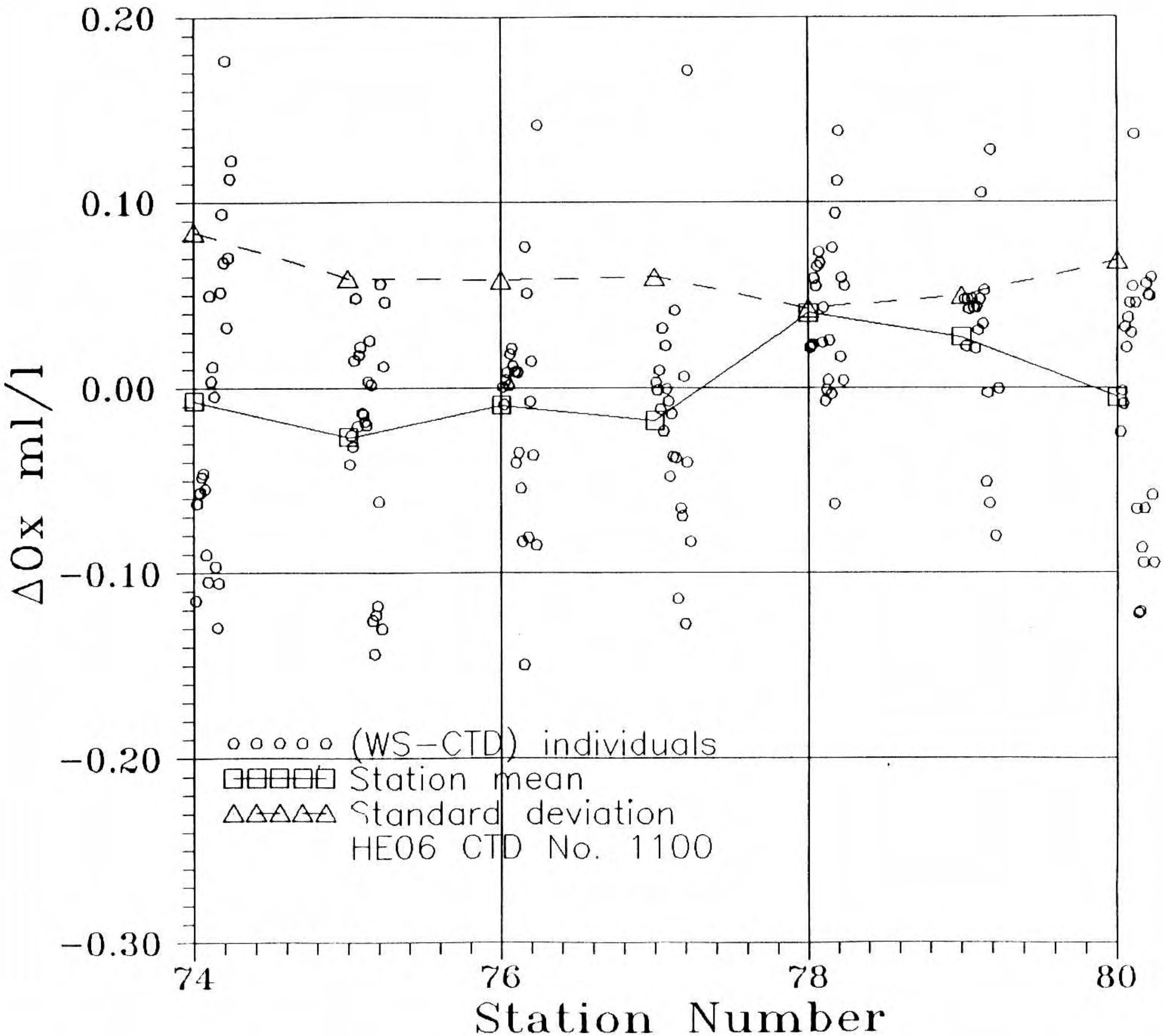


0.006
0.004
0.002
−0.000
−0.002
−0.004
−0.006
−0.008
ΔC mS/cm
0
2000
4000
6000
Pressure dbars
HE06 CTD No. 1100
Stations 74–80



Worth-Met NA
HE06D076 CTD #1100
HE06D074 CTD # 1100
HE06D075 CTD # 1100
AT19D208 WUNSCH
HE06D077 CTD # 1100
AT19D202 WUNSCH
HE06 WS
3.40
2.90
2.40
1.90
1.40
34.850
34.900
34.950
35.000
Salinity psu
Potential Temperature °C


0.20
0.10
0.00
−0.10
−0.20
−0.30
ΔOx ml/l
0
2000
4000
6000
Pressure dbars
HE06 CTD No. 1100
Stations 74–80


0.20
0.10
0.00
−0.10
−0.20
−0.30
ΔOx ml/l
74
76
78
80
Station Number
○○○○○ (WS−CTD) individuals
Station mean
△△△△△ Standard deviation
HE06 CTD No. 1100

References

Allen, J., 1992. 1 WHOI CTD MicroVAX II Data Acquisition System Part 1, 32 pages: Installation Manual. 2 WHOI CTD MicroVAX II Data Acquisition System Part 2, 65 pages: Operator's Guide. 3 WHOI CTD MicroVAX II Data Acquisition System Part 3, 38 pages: Reference Manual. 4 WHOI CTD MicroVAX II Data Acquisition System Part 4, 21 pages: Guide to Writing Programs to Accessing the Global Section.

Benson, B. and D. Krause, Jr., 1984. The concentration and isotopic fractionation of oxygen dissolved in fresh and seawater in equilibrium with the atmosphere. *Limnology Oceanography* **29**(3), 620–632.

Brown, N. L., and G. K. Morrison, 1978. W.H.O.I./Brown conductivity, temperature and depth microprofiler. Woods Hole Oceanographic Institution Technical Report No. 78-23, pp. 246.

Carritt, D. E., and Carpenter, J. H., 1966. Comparison and evaluation of currently employed modifications of the Winkler method for determining dissolved oxygen in seawater; A NASCO Report., *Sears Foundation: Journal of Marine Research*, 24(3), 286–318.

Enns, T., P. F. Scholander, and E. D. Bradshaw, 1965. Effects of hydrostatic pressure on gases dissolved in water. *The Journal of Physical Chemistry*, **69**(2), 389–392.

Greene, M. W., R. D. Gafford and D. G. Rohrbaugh, 1970. A Continuous-Profiling Deep-Submersible Dissolved Oxygen Monitor. *Marine Technology Society. Six Annual Reprints*, **2**, 1502 pp.

Fofonoff, N. P., 1985. Physical properties of seawater: A new salinity scale and equation of state for seawater. *Journal of Geophysical Research, 90*, 3332–3342.

Fofonoff, N. P., S. P. Hayes, and R. C. Millard, 1974. WHOI/Brown CTD microprofiler: methods of calibration and data handling. Woods Hole Oceanographic Institution Technical Report No. 74-89, pp. 66.

Fofonoff, N. P., and R. C. Millard, Jr., 1983. Algorithms for computation of fundamental properties of seawater. UNESCO Technical Papers in Marine Science 44.

Fofonoff, N. P., and R. C. Millard, 1991. Calculation of physical properties of seawater. In: WOCE operations manual, T. Joyce, editor, WOCE Hydrographic Programme Office, WHPO 91-1, Woods Hole Oceanographic Institution, pp.

Joyce T. (Editor) 1991. WOCE Operations Manual, Sections Titled: CTD Oxygen Calbration Procedure by R. Millard, and Calculation of Physical Properties of Seawater N. P. Fofonoff and R. C. Millard, WHPO 91-1.

Joyce, T., C. Corry, and M. Stalcup, editor, 1991. *Requirements for WOCE Hydrographic Programme Data Reporting*, WHPO-90-1, Review 1, WOCE Report No. 67/91.WHP Office, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts, 70 pp.

JPOTS Editorial Panel (H. D. Dooley, O. I. Mamayev, R. C. Millard, and K. Tiara), 1991. Processing of Oceanographic Station Data. UNESCO, 7, place de Fontenoy, 75700, Paris, 138 pp.

Kanwisher, J., 1959. Polarographic oxygen electrode. *Limnology Oceanography*, **4**(2), 210–217.

Knapp, G. P. and M. C. Stalcup, 1987. Progress in the Measurement of Salinity and Oxygen at Woods Hole Oceanographic Institution. Woods Hole Oceanographic Institution Technical Report No. 87–4. 27 pages.

Knapp, G. P., M. C. Stalcup and R. J. Stanley, 1990. Automated Oxygen Titration and Salinity Determination. Woods Hole Oceanographic Institiution. Technical Report No. 90–35, 25 pages.

Mangum, B. W., 1990. A New Temperature Scale, The International Temperature Scale of 1990, Is Adopted. NCSL newsletter. 10 p.

Millard, R. C., Jr., 1982. CTD calibration and data processing techniques at WHOI using the 1978 practical salinity scale. Marine Technology Society Conference paper.

Millard, R. C., 1986. International Oceanographic Tables Volume 4. Properties derived from the equation of state of seawater, 1980. UNESCO Technical Papers in Marine Science No. 40, 108 pp.

Millard, R., 1991. CTD oxygen calibration procedure. In: WOCE Operations Manual, T. Joyce, editor, WOCE Hydrographic Programme Office, WHOI 91-1, Woods Hole Oceanographic Institution.

Millard, R. C., A. Blumer, and N. Galbraith, 1978. A digital tape format for Woods Hole Oceanographic Institution CTD data. WHOI Technical Report No. 78–43.

Millard, R., G. Bond, and J. Toole, 1992. Implementation of a titanium strain gauge pressure transducer for CTD applications. *Deep-Sea Research*, 1009–1021.

Millard, R. C., and N. Galbraith, 1982. WHOI processed CTD data organization. WHOI Technical Report No. 82–37, 36 pages.

Millard, R. C., Jr., B. J. Lake, N. L. Brown, J. M. Toole, D. Schaaf, K. Yang, H. Yu and L. Zhou, 1990. US/PRC CTD Intercalibration Report 1986–1990. WHOI Technical Report 90-53.

Millard, R. C., W. B. Owens, and N. P. Fofonoff, 1990. On the calculation of the Brunt-Väisälä frequency. *Deep-Sea Research*. **37**, 167–181.

Millard, R., J. Toole and M. Swartz, 1980. A fast responding temperature measurement system for CTD applications. *Ocean Engineering*, 7, 413–427.

Owens, W. B., and R. C. Millard, Jr., 1985. A new algorithm for CTD oxygen calibration. *Journal of Physical Oceanography*, **15**, 621–631.

Perkins, R. G., and Lewis, E. L., 1980. The practical salinity scale 1978: fitting the data. *IEEE J. Oceanic Eng., OE-5*, (1), 9–16.

Press, W., B. Flannery, S. Teukolsky, and W. Vetterling, 1986. *Numerical Recipes: The Art of Scientific Computing*, Cambridge University Press, 818 p.

Rusby, R. L. 1989. Adoption of the International Temperature Scale of 1990, IPS-90. National Physical Laboratory, Teddington, Middlesex, UK. 5 p.

Saunders, P. M., 1990. The international temperature scale of 1990. ITS-90. *WOCE Newsletter*, **10**, IOS, Wormley, U.K.

Stalcup, M., R. Key, A. Mantyla, R. Millard, Contributors, 1990. *Requirement's for WOCE Hydrographic Programme Data Reporting*, WHPO-90-1, WHP Office, Dept. of Physical Oceanography, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts, 27 pp.

UNESCO, 1986. Progress on Oceanographic tables and standards 1983-1986: Work and recommendations of UNESCO/SCOR/ICES/IAPSO Joint Panel. UNESCO Technical Papers in Marine Science, No. 50, 59 pp.

UNESCO, 1988. The acquisition, calibration, and analysis of CTD data. UNESCO Technical Report 54, 92 pp.

Weiss, R. F., 1970. The solubility of nitrogen, oxygen, and argon in water and seawater. *Deep-Sea Reseach*, **17**, 721–735.

Worthington, L. V. and W. Wright, 1970. The Water Masses of the North Atlantic Ocean. A Volumetric Census of Temperature and Salinity. American Geographical Society. 12 pp.

UNOLS RVTEC

TECHNICAL PERSONNEL LISTING

First edition

Updated: 20 September 1993

CONTENTS

INTRODUCTION AND SUGGESTED FORMAT

This catalog represents an initial output from a developing database of equipment and talents available at the technical support departments of the various UNOLS institutions. It is hoped distribution of this information will give users more information on the strengths of particular institutions, and will better allow institutions to assist each other with exchange of talents and equipment in both emergency and nonemergency situations.

A suggested format for submitted data is given below. It is not intended to be a straitjacket, and submission in alternate formats is acceptable. Feel free to add or omit specific categories as you desire.

SUGGESTED BIOGRAPHICAL STRUCTURE WITH EXAMPLES

1. Name (let's call our example person Joe Wirehead).

2. Job Title.

3. Group or assignment. (Shipboard techs, geological, etc.).

4. Institution, mailing address, phone, FAX, EMAIL.
   (This is how you contact Joe to ask him a question).

5. Scheduling supervisor, with address, phone, FAX, EMAIL.
   (This is the person you contact to request Joe's services, for example, presence on a cruise).

6. Education & degrees.
   (Include short courses: i.e. ADCP training courses).

7. Citizenship and passport status.

8. Certifications and Licenses.
   (SCUBA, Amateur or Commercial Radio operations, dive officer, EMT, Paramedic, radiation safety officer, vessel captain, etc.)

9. Proficiency- equipment and instruments.

WORKING DEFINITION: proficiency on an instrument or system is defined here as ability to set up, verify proper function, operate, troubleshoot, do field repair, and train others in the use of the instrument. In simple terms, is this person competent to take charge of a system on a 2 week cruise if he/she is the only person on the boat who has ever seen the equipment before?

For this section, include actual models of instruments when possible (e.g Neil Brown Mk III, Aanderaa RCM-4). Obviously expertise can carry over from one device to another, but knowing specific models can help (for example) when someone has a problem with a particular instrument.

Suggested format: instrument type followed by Models, e.g:
CTD's: Neil Brown Smart, MK III; Applied Microsystems STD-12, Sea Bird 25.

Some suggested general categories:
Acoustic releases.
ADCP's.
Biological samplers (Mocness, etc.)
Coring equipment.
Current meters.
CTD's.
Dataloggers.
Drifters.
Echosounders.
Fluorometers.
Grabs.
Moorings (deployment and retrieval of)
Multiparameter analyzers.
Nets (plankton, trawls).
Navigation equipment: transit, GPS, LORAN.
Nutrient analyzers.
Radio communications equipment: shortwave, HF, VHF, SatCom.
Rosette sampler.
ROV's.
Salinometers, field and laboratory.
Seabeam.
Sidescan sonar.
Sediment traps.
Subbottom profilers.
Thermosalinographs.
Transmissometers.
Weather equipment & buoys.

Additional categories are solicited.

11. Proficiency - computers, software, communications.

Proficiency definitions and format as above.

Suggested categories:
Hardware (e.g. IBM-PC, Mac II, VAX, SUN, Onset Tattletale).
Operating systems (MS-DOS, Mac, Unix, VMS, Windows, OS/2, etc.)
Applications (Lotus, Quattro, WordPerfect, CAD, DBase).
Programming Languages (BASIC, C, Pascal, FORTRAN, also types of applications written).
Communications & interfaces (A/D, D/A, RS-232, RS-422, NMEA-0183, SAIL, 4-20 mA, Ethernet, etc.).

12. Other skills.
List anything here that is germane and not obvious from other sections:
Calibration analysis.
Carpentry, cabinetmaking.
Electronics design.
Foreign languages spoken.
Metalworking.
Mechanical design.
Mooring design.
Radiochemistry.
Ropecraft.
Small boat handling.
Welding.
Wet chemistry (Winkler DO analysis, etc.)

13. Experience.
Years at present position.
Total years of experience in oceanography.
Years of experience in other germane fields.

Vessel experience: list of vessels worked on.

Cruise experience: list of cruises/projects & years worked on, responsibilities.

14. Potential availability, ability to travel, & current cost/day.
Example: "Potentially available for cruises in 93-94 up to 3 weeks duration. Will travel worldwide. 1993 charge: $300.00 per day, 7 days per week, charged home institution to home institution, plus travel expenses and perdiem."

15. References with name, address, phone, FAX, EMAIL. This should ideally be one or more science people who are not in the tech's chain of command and who can provide unbiased critical evaluation on the person's performance at sea.

RESEARCH VESSEL TECHNICAL ENHANCEMENT COMMITTEE

TECHNICIAN EXCHANGE POLICY STATEMENT

"In order to achieve the stated purpose of fostering activities that will lead to enhanced technical support for sea-going scientific programs, RVTEC recommends that all member institutions develop programs for training and skills enhancement of technical personnel. Each program should include at a minimum a statement of purpose, identify mechanisms for funding and fields of interest in which training will be focused, and provide time tables for personnel training.

Each program shall be structured around the particular fields of interest inherent to that institution's scientific goals and programs, and strive to develop expertise critical to the advancement of oceanographic science.

RVTEC shall act in an advisory capability in the development of each institution's program and review the final policy for redundancy of training efforts within RVTEC institutions, or programs that are not within the stated purpose of RVTEC."

Newsgroups: sci.data.formats,news.answers,sci.answers
Path: news.miami.edu!wupost!cs.utexas.edu!asuvax!ncar!kiowa.scd.ucar.edu!ilana
From: ilana@kiowa.scd.ucar.edu (Ilana Stern)
Subject: Scientific Data Format Information FAQ
Message-ID: <1993Sep15.110012.10318@ncar.ucar.edu>
Followup-To: sci.data.formats
Summary: Where to find information on scientific data formats
Sender: news@ncar.ucar.edu (USENET Maintenance)
Reply-To: ilana@ncar.ucar.edu
Organization: NCAR/UCAR
Date: Wed, 15 Sep 1993 11:00:12 GMT
Approved: news-answers-request@MIT.Edu
Expires: Wed, 29 Sep 1993 07:00:00 GMT
Lines: 276
Xref: news.miami.edu sci.data.formats:252 news.answers:12581 sci.answers:433

Archive-name: sci-data-formats
Last-modified: 14 Sep 1993

This is the FAQ for the sci.data.formats newsgroup. Contents:

Each (major) section has a "Subject:" line, so you can search on the subject title above to find the section quickly.

Comments, corrections, or additions should be sent to Ilana Stern at ilana@ncar.ucar.edu.

Subject: How to use this document

Most FAQ (Frequently Asked Questions) documents list many questions and their answers. This FAQ is (mostly) devoted to answering only one question:

"Where can I find documentation and software for <X> data format?"

As the amount of information available over the networks has been increasing, so have the methods by which this information can be obtained. No longer is direct usage of FTP the only, or even the most frequent, method of obtaining data; we now have Gopher, Wais, and WWW, as well as many site-specific interfaces. Because the information itself may be accessible in many different ways, this FAQ will identify resources in terms of URLs (Uniform Resource Locators). This will also help us convert this FAQ to a hypertext document, so that it can be used with a WWW browser to go directly to any of the listed sources.

Here's a glossary, so you can decode the URLs if necessary to reach the sites:

ftp://host.name.domain/directory/[subdirectory/[filename]]	ftp site
http://host.name.domain/directory/[subdirectory/[filename]]	www server
telnet://host.name.domain	telnet site
gopher://host.name.domain	gopher server
wais://host.name.domain	wais server
news:newsgroup.name	newsgroup

So, for example, if a document is available at ftp://ncardata.ucar.edu/ it means that you should ftp to ncardata.ucar.edu, and the information is

in the top-level directory.
If you don't know what these information retrieval methods are, see the section "How to use the data retrieval methods".

Subject: How to get a current copy of this document

If you are reading this document after 29 Sep 1993, you are reading an outdated copy. A current copy of this document can be obtained by anonymous FTP to ftp://rtfm.mit.edu/pub/usenet/news.answers/data-formats.
If you don't know what FTP is, see the section "How to use the data retrieval methods".

If you can't use FTP, send email to mail-server@rtfm.mit.edu with

```
        send /pub/usenet/news.answers/data-formats
```

as the only text in the message (leave the subject blank).

A current hypertext version of this document can be obtained from http://fits.cv.nrao.edu/traffic/scidataformats/faq.html or from http://www.ncsa.uiuc.edu/SDG/Software/HDF/SciDataFormatsFAQ.html. If you would like to archive this FAQ in either hypertext or plaintext format, and want to receive a new copy automatically at every update, please send me email.

Subject: Resources for format information

1) CDF
2) FITS
3) GRIB
4) HDF
5) netCDF
6) Miscellaneous graphics formats

1. CDF
CDF (Common Data Format) is a library and toolkit for multi-dimensional data sets. The basic component of CDF is a software programming interface that is a device independent view of the CDF data model. A user's guide and software is available from ftp://nssdca.gsfc.nasa.gov/cdf/.
Questions can be directed to cdfsupport@nssdca.gsfc.nasa.gov.

2. FITS

FITS (Flexible Image Transport System) is the standard data interchange and archival format of the worldwide astronomy community. The NOST Standard and User's Guide, some software, and test files are available from ftp://nssdca.gsfc.nasa.gov/fits/.
The site ftp://fits.cv.nrao.edu/fits (accessible via WWW at http://fits.cv.nrao.edu) has other software and a different set of test files, and electronic copies of FITS proposals that are under development or in the international approval process. Archives of sci.data.formats that are of interest to astronomers are here, too.
A registered WAIS index that can be searched for FITS information is at http://info.cern.ch:8001/fits.cv.nrao.edu:210/nrao-fits.
The newsgroup news:sci.astro.fits is devoted to discussion of the FITS format.
If you've searched all these resources and still have questions, you can direct them to fits@nssdca.gsfc.nasa.gov.

3. GRIB

GRIB (GRid In Binary) is the World Meteorological Organization (WMO) standard for gridded meteorological data. Unfortunately it is still not very "standard", as some organizations use their own versions. A format description for WMO GRIB can be found at
ftp://ncardata.ucar.edu/datasets/ds084.5/format.grib

and read code is in the file access_grib.f in the same directory.
If you need GRIB to read ECMWF data, the above format description, along with the ECMWF-specific parameter table, and a list of differences between WMO and ECMWF GRIB, is in ftp://ncardata.ucar.edu/datasets/ds111.2/format. Read code can be found in ftp://ncardata.ucar.edu/datasets/ds111.2/software.
If all else fails, contact Ilana Stern at ilana@ncar.ucar.edu.

4. HDF

HDF (Hierarchical Data Format) is a self-defining file format for transfer of various types of data between different machines. The HDF library contains interfaces for storing and retrieving compressed or uncompressed raster images with palettes, and an interface for storing and retrieving n-Dimensional scientific datasets together with information about the data, such as labels, units, formats, and scales for all dimensions.
Source code and documentation are on ftp://ftp.ncsa.uiuc.edu/HDF. Some general information on HDF, including a FAQ, is available from http://www.ncsa.uiuc.edu/SDG/Software/HDF/HDFIntro.html.

5. netCDF

NetCDF (Network Common Data Form) is an interface for scientific data access which implements a machine-independent, self-describing, extendible file format. Source code and documentation for the netCDF data access library is available from ftp://ftp.unidata.ucar.edu/pub/netcdf or gopher://gopher.unidata.ucar.edu:70. A FAQ is here also, as well as pointers to other available software packages which use netCDF data. Questions and answers about netCDF can be searched in the index wais://wais.unidata.ucar.edu:210/unidata-support-netcdf.src.
A recent paper (Jenter and Signell, 1992) which provides a good introduction to netCDF is available as ftp://crusty.er.usgs.gov/pub/netcdf.asce.ps.
A mailing list, netcdfgroup@unidata.ucar.edu, exists for discussion of the netCDF interface, and for announcements of netCDF news: to subscribe, send a request to netcdfgroup-adm@unidata.ucar.edu. The archives of netcdfgroup are available from ftp://ftp.unidata.ucar.edu/mail-archives/netcdfgroup, and also in the searchable index wais://wais.unidata.ucar.edu:210/netcdf-group.src.
For more information, contact support@unidata.ucar.edu.

6. Miscellaneous graphics formats

These formats for storing graphics files -- TIFF, GIF, JPEG, FLI, CGM, and so on -- are more properly discussed in news:comp.graphics. A small amount of documentation on these and other graphics formats is on ftp://zamenhof.cs.rice.edu/pub/graphics.formats; other archive sites are ftp://ftp.ncsa.uiuc.edu/misc/file.formats/graphics.formats, and ftp://telva.ccu.uniovi.es/pub/graphics/Image.
The comp.graphics FAQ and resource file have more information on where to find read and conversion programs for these formats. You can find them at ftp://rtfm.mit.edu/pub/usenet/news.answers.
A good (hardcopy) reference for graphics formats is _Graphics File Formats_, by David C. Kay and John R. Levine (Windcrest Books, ISBN 0-8306-3060-0, about US$30.00 in paperback).

Subject: How to use the data retrieval methods

This section only describes FTP and telnet in any detail; for other methods, FTP sites are given, so you can get information on them yourself.

1) How to use FTP
2) How to use telnet
3) Gopher information
4) Wais information
5) WWW information

1. How to use FTP

FTP (File Transfer Protocol) allows transfer of files between two computers which are on the Internet. To access the FTP areas listed here, at your system prompt type "ftp" followed by the name of the desired system. For example, to access ncardata.ucar.edu you'd type

```
ftp ncardata.ucar.edu
```

Use "anonymous" as your login and your email address as the password (if requested).

[Note: quotes ("like this") are used to set off names of directories and files, or commands you'd type, and are not part of these names.]

Not all FTP systems accept the same commands, but here's a list of the most useful:

```
ls      list files in the current directory.
cd      change directory, e.g. "cd wx" changes to the wx directory.
binary  sets binary mode
ascii   sets ascii mode (the default).  Use for retrieving text.
get     retrieves a file, e.g. "get readme" gets a file called readme.
bye     exits FTP.
```

If you can't seem to connect to the site, check to see if it is a telnet site. If it is, follow the instructions in the following section instead.

If you can't FTP from your site, use one of the following ftp-by-mail servers:

```
ftpmail@decwrl.dec.com
ftpmail@src.doc.ic.ac.uk
ftpmail@cs.uow.edu.au
ftpmail@grasp.insa-lyon.fr
```

Send an e-mail message to the closest address, with the lines:

```
reply your_address@some.where       <- with your email address
connect ncardata.ucar.edu           <- for example
cd datasets/ds111.2/software
get access_sun.f
quit
```

For complete instructions, send a one-line message reading "help" to the server. Please don't ask me for help!

2. How to use telnet

Type "telnet" followed by the name or IP number of the desired system. These publicly accessible systems generally allow you to log in but put you in a restricted shell, from which only a certain menu of commands is available. The description for the site will include the login to use.

If you can't seem to connect to the site, re-check its description in the document; if it's an FTP site, follow the instructions in the previous section instead.

3. Gopher information

Available by ftp at ftp://rtfm.mit.edu/pub/usenet/news.answers/gopher-faq.

4. Wais information

Available by ftp at ftp://rtfm.mit.edu/pub/usenet/news.answers/wais-faq/getting-started.

5. WWW information

Available by ftp at ftp://rtfm.mit.edu/pub/usenet/news.answers/www/faq. WWW is so easy to use that you might as well just hop in and try it, so ask your sysadmin if you have a WWW browser such as NCSA Mosaic.

Subject: Why isn't my favorite format on this list?

If you don't see a format you're interested in here, it could be one of three reasons. First of all, there are a lot of formats which are out of the scope of this newsgroup: it ain't named *sci*.data.formats for nuthin', you know. Formats used in commercial spreadsheet and word-processing software aren't scientific data formats, and aren't discussed in this group.

Second, it may be that nobody has given the FAQ organizer any information on sources for information on that format. So ask the newsgroup -- and if you do get a response, please let me know what it is!

Finally, you may ask on the net, and hear nothing, because the data format description just *isn't* publicly available. For most scientific data formats, this is a Bad Thing, and most archivists and scientists *want* to have their format information available. If you have such information, but don't have resources to make it available, please ask around and see if you can get it into an FTP area or other resource. Please don't publicize private or proprietary formats without the permission of the author, though.

Subject: netCDF FAQ

Frequently Asked Questions About netCDF

This article contains answers to some of the most frequently asked netCDF questions on the netcdfgroup mailing list and in the email sent to support@unidata.ucar.edu. The list is archived in pub/netcdf/FAQ on ftp.unidata.ucar.edu, and is currently maintained by Russ Rew (russ@unidata.ucar.edu).

List of Questions:

1: What Is netCDF?

A: NetCDF (network Common Data Form) is an interface for scientific data access and a freely-distributed software library that provides an implementation of the interface. It was developed by Glenn Davis, Russ Rew, and Steve Emmerson at the Unidata Program Center in Boulder, Colorado. The netCDF library also defines a machine-independent format for representing scientific data. Together, the interface, library, and format support the creation, access, and sharing of scientific data.

netCDF data is:

- Self-Describing. A netCDF file includes information about the data it contains.

- Network-transparent. A netCDF file is represented in a form that can be accessed by computers with different ways of storing integers, characters, and floating-point numbers.

- Direct-access. A small subset of a large dataset may be accessed efficiently, without first reading through all the preceding data.

- Appendable. Data can be appended to a netCDF dataset along one dimension without copying the dataset or redefining its structure. The structure of a netCDF dataset can be changed, though this sometimes causes the dataset to be copied.

- Sharable. One writer and multiple readers may simultaneously access the same netCDF file.

2: How do I get the netCDF software package?

A: Via anonymous FTP from

host: ftp.unidata.ucar.edu
file: pub/netcdf/netcdf.tar.Z

Make sure the file is transmitted in BINARY mode. This is version 2.3.2, last updated in June 1993. In addition there are currently two patch files (2.3.2-patch1 and 2.3.2-patch2) available from the same directory that fix bugs found so far in version 2.3.2.

Q: What does netCDF run on?

A: Version 2.3.2 of netCDF has been tested on the following platforms:

CRAY Y-MP	UNICOS 6.1.6
DEC Alpha	OSF/1 1.2
DEC VAX	VMS 5.5-2
DEC VAX	Ultrix 4.3
DECstation 3000	Ultrix 4.3
HP-9000/7xx	HPUX 9.0
IBM PS/2	MSDOS 5.0
IBM PS/2	OS/2 1.2
IBM RS-6000	AIX 3.2
NeXT	NeXTOS 3.0
SGI Iris	IRIX 4.0.5F
SPARCstation	Solaris 2.1
SPARCstation	SunOS 4.1.3

The experience of outside developers indicates that netCDF is relatively easy to port to any system that has a C compiler and an XDR library (eXternal Data Representation, used for NFS and usually supplied by vendors).

4: What has changed since the netCDF 2.2 release in December 1991?

A: NetCDF 2.3, made available in April 1993, contains bug fixes, portability enhancements, performance enhancements, and a few new capabilities.

Some new optimizations for the library result in significant speedups for accessing cross-sections involving non-contiguous data.

New capabilities include some additional interfaces that provide a more general form of hyperslab access. This supports sub-sampling along specified dimensions and a mapping between the points of the hyperslab and the memory locations of the corresponding values. In a generalized hyperslab, an index mapping vector is used to define the mapping between points in the generalized hyperslab and the memory locations of the corresponding values, so data values that are written or read need no longer be contiguous in memory.

There are also some new interfaces that can be used to write, read, and inquire about records, where a record may contain multiple variables of different types and shapes. Where before you had to access a record's worth of data using multiple calls, now you will be able to use a single call.

The ncdump utility supports several new command-line options including the ability to specify for which variables data values will be output, to provide brief annotations in the form of CDL comments to identify data values for large multidimensional variables, or to provide full annotations in the form of trailing CDL comments for every data value.

The current release also includes a prototype implementation of a C++ interface for the netCDF data access library. It provides all the functionality of the previous C interface, improves type safety by eliminating use of void* pointers, and is somewhat simpler to use than the C interface. With the C++ interface, no IDs are needed for netCDF and less indirection is required for dealing with dimensions. However, since this is a prototype interface and implementation, it may be changed before a supported version is released.

5: What is the connection between netCDF and CDF?

A: CDF was developed at the NASA Space Science Data Center at Goddard, and is freely available. It was originally a VMS FORTRAN interface for scientific data access. Unidata reimplemented the library from scratch to use XDR for a machine-independent representation, designed the CDL text representation for netCDF data, wrote a User's Guide and made other additions including aggregate data access, single-file implementation, named dimensions, and variable-specific attributes.

NetCDF and CDF have evolved independently. CDF now supports many of the same features as netCDF (aggregate data access, XDR representation, single-file representation, variable-specific attributes), but some differences remain (netCDF doesn't support native-mode representation, CDF doesn't support named dimensions). There is no compatibility between data in CDF and netCDF form, and as yet no translation software exists to convert data in one form to data in the other form.

6: What is the connection between netCDF and HDF?

A: The National Center for Supercomputing Applications (NCSA) developed the HDF software and makes it freely available. HDF is an extensible data format for self-describing files that was developed independently of netCDF. Applications and utilities based on HDF are available that support raster-image manipulation and display and browsing through multidimensional scientific data. The HDF software includes a package of routines for accessing each HDF data type, as well as a lower-level interface for building packages to support new types. HDF supports both C and Fortran interfaces, and it has been successfully ported to a wide variety of machine architectures and operating systems. HDF emphasizes a single common format for data, on which many interfaces can be built.

NCSA has implemented software that provides a netCDF interface to HDF. With this software, it is possible to use the netCDF calling interface to place data into an HDF file. The netCDF calling interface has not changed and netCDF files stored in XDR format are readable, so existing programs and data will still be usable (although programs will need to be relinked to the new library). There is currently no support for the mixing of HDF and netCDF structures. For example, a raster image can exist in the same file as a netCDF object, but you have to use the Raster Image interface to read the image and the netCDF interface to read the netCDF object. The other HDF interfaces are currently being modified to allow multi-file access, closer integration with the netCDF interface will probably be delayed until the end of that project.

Eventually, it may be possible to integrate netCDF objects with the rest of the HDF tool suite. Such an integration will then allow tools written for netCDF and tools written for HDF to both interact intelligently with the new data files.

7: Is netCDF available for the Apple Macintosh?

A: Unidata doesn't test or maintain a version of netCDF for Macintoshes, but see the files in the directory pub/netcdf/mac/ on ftp.unidata.ucar.edu for some contributed notes and Macintosh MPW makefiles for porting netCDF to an Apple Macintosh.

8: What are some references to netCDF?

A: Hard-copies of some of these are available from the Unidata Program Center, P.O. Box 3000, Boulder, CO 80307-3000:

Rew, R. K., G. P. Davis, and S. Emmerson, NetCDF User's Guide, An Interface for Data Access, Version 2.3, April 1993. (Available from Unidata or by anonymous FTP from ftp.unidata.ucar.edu in the file pub/netcdf/guide.ps.Z)

Rew, R. K. and G. P. Davis, "NetCDF: An Interface for Scientific Data Access," IEEE Computer Graphics and Applications, Vol. 10, No. 4, pp. 76-82, July 1990.

Rew, R. K. and G. P. Davis, "The Unidata netCDF: Software for Scientific Data Access," Sixth International Conference on Interactive Information and Processing Systems for Meteorology, Oceanography, and Hydrology, Anaheim, California, American Meteorology Society, February 1990.

Jenter, H. L. and R. P. Signell, 1992. "NetCDF: A Freely-Available Software-Solution to Data-Access Problems for Numerical Modelers". Proceedings of the American Society of Civil Engineers Conference on Estuarine and Coastal Modeling. Tampa, Florida. (Also available via anonymous FTP from sparky.er.usgs.gov in the file pub/netcdf.asce.ps)

Fulker, D. W., "Unidata Strawman for Storing Earth-Referencing Data," Seventh International Conference on Interactive Information and Processing Systems for Meteorology, Oceanography, and Hydrology, New Orleans, La., American Meteorology Society, January 1991.

Brown, S. A, M. Folk, G. Goucher, and R. Rew, "Software for Portable Scientific Data Management," Computers in Physics, American Institute of Physics, Vol. 7, No. 3, May/June 1993, pp. 304-308.

9: What are the files in pub/netcdf/ on ftp.unidata.ucar.edu?

A:

File	Purpose
README	general information about netCDF
FAQ	this file of frequently asked questions
utilities.txt	a list of software packages currently available or under development for manipulating and displaying netCDF data
guide.ps.Z	a compressed PostScript file of the NetCDF User's Guide. This is included in the netcdf.tar.Z distribution, so you don't need both.
ncprogs.ps	a draft PostScript document describing an initial set of netCDF operator and utility programs under development
ncprogs.txt	an ASCII version of ncprogs.ps
Conventions/	a directory intended for documents describing discipline-specific conventions from various groups that are using netCDF.
cdl/	a directory containing some examples of CDL files (an ASCII representation for netCDF files).
msdos/	a directory containing executables and binaries fc netCDF under MSDOS 5.0. These can also be built from the sources in netcdf.tar.Z, if you have the necessary Microsoft compilers.
mac/	a directory containing notes and Macintosh MPW makefiles for porting netCDF to an Apple

10: Why do I get XDR error messages when trying to write data?

A: Probably due to a write error, perhaps because of exceeding disk quotas, a full device, or permission problems. The netCDF library is usually built on a vendor-supplied XDR library layer. Sometimes errors occur in the XDR layer rather than the netCDF layer, and in this case the error messages from the XDR layer can be cryptic. For example, if the user is writing data and exceeds a disk space quota, this might be detected first in the XDR layer. Although we try to catch and elaborate on errors that occur in the XDR layer, sometimes there is not enough information passed up to calling routines from the XDR layer to make construction of a meaningful error message possible.

11: What is the best way to represent [some particular data] using netCDF?

A: There are many ways to represent the same information in any general-purpose data model. Choices left up to the user in the case of netCDF include which information to represent as variables or as variable attributes; what names to choose for variables, dimensions, and attributes; what order to use for the dimensions of multidimensional variables; what variables to include in the same netCDF file; and how to use variable attributes to capture the structure and meaning of data. We provide some guidelines in the NetCDF User's Guide (e.g Section 2.3.2, ''Differences between Attributes and Variables''), but we've found that a little experience helps. Occasionally we have decided it was useful to change the structure of netCDF files after experience with how the data is used.

12: Are there plans to add facilities for data compression to netCDF?

A: We have no plans to add data compression to netCDF (although we do hope to eventually add a form of transparent data packing on write and unpacking on read whenever the reserved attributes "_Nbits", "_Scale", and "_Offset" are defined).

Hyperslab access and direct access to individual array values conflict with most simple compression schemes. With netCDF, the elements of an array variable can be filled in any order or as cross-sections in any direction. NetCDF permits writing elements in one order and reading them later in different orders.

Some compression methods require that all the data to be compressed are known before starting the compression. Techniques like run-length encoding or anything that depends on exploiting similarities in nearby values can't be used if nearby values aren't all known at the time some of the data are to be written.

An alternative that can be implemented above the netCDF library is to adopt a convention for compressed data that uses a "compression" attribute to encode the method of compression, e.g.

```
x:compression = "rle" ;
```

for run-length encoding of the data in a variable x. Then when you write the data, compress them into a bland array of bytes and write all the bytes. Note that it would be difficult to define the size of such a variable in advance, since its compressed size depends on its values. You would also have to give up on hyperslab access for such variables, but instead read the compressed array in all at once and uncompress it before using it.

13: What utilities are available for netCDF?

A: The only utilities available in the current netCDF distribution from Unidata are ncdump and ncgen, for converting netCDF files to an ASCII human-readable form and for converting from the ASCII human-readable form back to a binary netCDF file or a C or FORTRAN program for generating the netCDF file. In addition, the first release of the netcdf operators package, containing three general-purpose netCDF operators, is available from ftp.unidata.ucar.edu in the file ncopers.tar.Z. See the description of the file utilities.txt in the answer to the question above on important files in /pub/netcdf on ftp.unidata.ucar.edu for more information.

14: What is the status of implementing additional netCDF operators?

A: Work on additional netCDF operators has been suspended until we can get more programmer resources, due to a higher priority recently placed on development of an event-driven network data distribution system. Volunteers to implement one or more of the netCDF operators are hereby solicited. We will coordinate community efforts to avoid duplication of effort, so before you volunteer to work on one of the planned netCDF operators, please contact support@unidata.ucar.edu to find out if someone else is already working on it.

15: Is there a mailing list for netCDF discussions and questions?

A: Yes - there are two. The mailing list

netcdfgroup@unidata.ucar.edu

has over two hundred readers and gateways; you can subscribe or unsubscribe to the mailing list by sending mail to

netcdfgroup-adm@unidata.ucar.edu

If you would prefer to get only a single daily digest of the postings to the netcdfgroup mailing list, subscribe instead to the netcdfdigest mailing list by sending a request to

netcdfdigest-adm@unidata.ucar.edu

All the postings to the netcdfgroup mailing list are archived in the file mail-archives/netcdfgroup available via anonymous FTP from ftp.unidata.ucar.edu.

16: How do I make a bug report?

A: If you find a bug, send a description to

support@unidata.ucar.edu

This is also a better address to use for questions or discussions about netCDF that you think are not appropriate for the entire netcdfgroup mailing list.

17: Is there any WAIS or gopher access to netCDF information?

A: Yes, 'netcdf-group.src' is a WAIS source that provides a full-text search of the netcdfgroup mailing list archive and is registered with think.com for general use.

Another WAIS source, 'unidata-support-netcdf.src', provides a full-text search of the support questions and answers about netCDF provided by Unidata support staff.

Both of these WAIS servers and access to other information about Unidata can be accessed most easily through the Unidata gopher server at gopher.unidata.ucar.ed

18: How widely is netCDF used?

A: The netcdfgroup and netcdfdigest mailing lists have 290 addresses (some of which are aliases to more addresses) in fifteen countries. Several groups have adopted netCDF as a standard way to represent some forms of scientific data.

The global ocean modeling effort at Los Alamos National Laboratory (LANL), as part of the DOE CHAMMP effort and one of the DOE Grand Challenges, has selected netCDF as the archival format for its computational data. An effort to bring netCDF up on the parallel disks on the CM-5 is planned to begin shortly.

The National Center for Supercomputing Applications has incorporated the netCDF 2.3 interfaces into the latest release of their HDF software, permitting HDF tools that use this interface to be applied to netCDF datasets that are either XDR- or HDF-encoded.

The Computer Planning Committee of NOAA's Pacific Marine Environmental Laboratory (PMEL) endorsed netCDF as the preferred data format for the Laboratory in early 1993. PMEL has developed the EPIC system for management, display and analysis of oceanographic time series and hydrographic data. EPIC utilizes netCDF as its primary data format. EPIC toolkits for netCDF include a data file I/O library, which is layered on top of the netCDF library, a netCDF calculator (nccalc) linked with a scientific graphics package (PPLUS), and a suite of display and analysis programs for oceanographic data.

Lamont-Doherty Earth Observatory of Columbia University has converted all Marine Geophysics data (gravity, magnetics and bathymetry) acquired in the past 40 years by scientists at L-DEO as well as at other institutions to netCDF. A package of programs to access, maintain and display those files has also been completed. The software package is available via FTP (pub/cdf/MGG_CDF.tar.Z on lamont.ldeo.columbia.edu)

The Sequoia 2000 project has adopted NetCDF as one of its supported formats. Sequoia 2000 is an Earth Science Database being developed at several University of California campuses and at other institutions around the state. Vizualization and analysis tools are under development that use NetCDF as an interchange format.

The Generic Mapping Tools (GMT), a Unix-based set of tools for data manipulation and display using PostScript, make use of netCDF for storage of 2-D gridded data sets. GMT is used worldwide by about 3000 scientists, according to the developers.

The Models-3 Project, being cooperatively pursued by the EPA's Atmospheric Research Laboratory and by North Carolina Supercomputing Center, is using an environmental-modeling-specific applications programming interface on top of UCAR's netCDF as the means for persistent storage of both observational and model-output data, as well as for storing sets of data-file-structure definitions and (prototype, so far) data-dependency graphs for scheduling the sets of programs which constitute their environmental models.

A group in the Atmospheric Chemistry Division at NCAR (the National Center for Atmospheric Research) that deals with UARS (Upper Atmospheric Research Satellite) data uses netCDF for their binary data format. Output from NCAR's High Altitude Observatory Division Thermospheric General Circulation Model (TGCM) and related models are converted to netCDF files for post-model visualization and diagnostic codes. NCAR's Research Aviation Facility will use netCDF to distribute all aircraft data, if performance tests on writing high rate data are satisfactory.

data archived and used in the "zeb" display and analysis system. Quick look data from various projects is distributed by RDP in netCDF. NetCDF is also the file format used by the (zeb-based) Integrated Sounding System.

PolyPaint is an interactive 3D visualization package from NCAR. PolyPaint V 3.4 uses netCDF for data sets, and storage of geometry information. PolyPaint+ alpha-version, being developed jointly by NCAR/MMM and LASP with funding from AISRP/NASA uses netCDF and DataHub from JPL.

The Cooperative Program for Operation Meteorology, Education, and Training (COMET), a program of UCAR, has created an extensive archive of meteorological case studies that contain observed and gridded data in netCDF. The netCDF definition in use was created by the Forecast Systems Laboratory (FSL), a NOAA agency, and files in this format are required for capatability with their PC-DARE workstations which are used in the COMET teaching classroom. COMET plans to create future case studies using the netCDF conventions currently being developed by a working group of Unidata, COMET, and FSL personnel.

The Southern Reginal Climate Center (A NOAA/NWS funded center) has decided to maintain all climate data in the netCDF. This includes the Zephyr/Unidata feed, ORNL-HCN/D data, NWS COOP data, and LAIS data. Currently, the Northeast Regional Climate Center is working on netCDF compatibility with a possible further view to conversion.

The Earth Scan Lab, an HRPT ground station at Coastal Studies Institute, is using both the Terascan TDF as well as the netCDF for ease of data exhange of AVHRR, TOVS and DCS data. Further, in conjunction with Woods Hole, Scripps and Texas A&M, CSI will be maintaining all oceanographic data in the netCDF.

DataHub from JPL, with funding from AISRP/NASA identifies and converts between a variety of data format, CDF, HDF, MMM/netCDF, FITS, PDS, ... Work is under way to support conversion from a variety of NASA data formats to netCDF used by the PolyPaint+ visualization system from NCAR's MMM division. (JPL Contact for DataHub: Tom Handley, thandley@spacemouse.jpl.nasa.gov)

A major component of the US Climate and Global Change program is the TOGA-TAO Array in the tropical Pacific, which proposes to maintain approximately 70 moored ATLAS wind and thermistor chain and current meter buoys, spanning the Pacific ocean from 95W to 137E in the equatorial wave guide. The TAO Project Office, at PMEL, has developed distribution and display software for the real-time data from the TAO buoys, in a point-and-click UNIX workstation environment. This software is distributed nationally and internationally. All data is stored and distributed in netCDF format. All graphics displays and animations are produced with the EPIC tools for working with netCDF data files.

The US Geological Survey's Branch of Atlantic Marine Geology uses netCDF to access a variety of scientific data sets, including output from circulation and sediment tranport models, sonar imagery, digital elevation models, and environmental sensor data. It is currently investigating the utility of using netCDF for data distribution on CDROM.

At the Woods Hole Oceanographic Institution, netCDF is used in several areas. Ships in the UNOLS fleet are recording measurements from the IMET systems in netCDF form. These data include wind, barometer, humidity, air and sea temperature, precipitation, short wave radiation, and GPS navigation. Data sets from these systems taken during the WOCE experiments in the Pacific have been archived recently at NCAR. Also, measurements from a diverse set of instruments deployed on buoys for the into netCDF form for processing and archival. Reports that describe the software systems used for these processing activities are available from WHOI.

Scripps Institution of Oceanography (SIO) and the University Corporation for Atmospheric Research (UCAR) conducted a multi-platform climate field project during March of this year based in Nadi, Fiji. All data from this experiment will be archived using Unidata's netCDF before release to the scientific community.

The Oregon State University Oceanographic research vessel WECOMA uses the netCDF library for primary scientific data logging. This includes navigational, meteorological, and other miscellaneous data. This logging is part of a client/server system for data distribution, display, and management known as XMIDAS. More details of the system are available from oce.orst.edu in the directory /pub/Wecoma, including a sample netCDF created data file from a Wecoma cruise.

NOAA's Forecast System Laboratories have adopted netCDF as a data access interface for some of their systems and applications.

A researcher at CSIRO Division of Oceanography in Australia is using netCDF for input and output files for a three dimensional coastal and estuarine hydrodynamic model.

Meteorological data from satellites is stored in netCDF form at CIRES (Cooperative Institute for Research in the Environmental Sciences) and several data analysis packages have been written to display and analyze the netCDF data.

A general purpose finite element data model (referred to as EXODUS II) utilizing netCDF has been developed at Sandia National Laboratories. It consists of a C and FORTRAN application programming interface (API) to read and write geometry and results (including time varying data) for finite element analyses. For more information, contact Larry Schoof (laschoo@somnet.sandia.gov).

NetCDF is the defacto standard for Analytical Data Interchange and Storage Standards (ADISS) for chromatography and mass spectrometry. NetCDF has been adopted by the Analytical Instrument Association (AIA), which includes all major analytical laboratory instrument vendors worldwide. ADISS supplies the Analytical Information Model on which the information content of these standards is based. The main uses of ADISS/netCDF by end users have been data communications and archival.

The AIA Standards Committee is currently developing an ADISS Data Dictionary for infrared spectroscopy. The AIA expects to use netCDF for implementing data interchange.

Another committee is developing an ADISS Data Dictionary for nuclear magnetic resonance and magnetic resonance imaging spectroscopy. After sufficient performance testing is completed, netCDF may be used for native storage for NMR, as well as data communications. Testing is expected to be completed at the end of summer 1993.

The Positron Imaging Laboratories and the Neuro-Imaging Laboratory of the Montreal Neurological Institute have selected netCDF as the data format for their medical image files. Conventions for variable and attribute names and values have been established for the medical imaging context. These conventions, along with a package of routines to assist in handling image data, make up the MINC (Medical Image NetCDF) format. For more information, contact Peter Neelin (neelin@pet.mni.mcgill.ca).

Several commercial analysis and data visualization packages have been

Visualization Data Explorer, Research Systems' IDL, and Wavefront's Data Visualizer.

Visualogic, Inc. (Bellevue WA, dave@zelda.nwra.com) is the developer of Dicer, a scientific data visualization tool for the Macintosh. Dicer recognizes and reads netCDF format files, in addition to HDF and generic binary array formats. Although Dicer is designed primarily for the visual inspection and analysis of volumetric data, it also provides data file translation, resampling, and reformating functions for netCDF and HDF files. Dicer is published and marketed by Spyglass, Inc., Savoy IL (217-355-6000).

SuperComputer Systems Engineering and Services Company (SSESCO) decided that for all future releases of their scientific visualization software, savi3D, netCDF will be the 'native' format. SSESCO has implemented a meta-file layer on top of the netCDF library, called MRAF. It handles multiple netCDF files as well as automatic max-min calculations, time-varying gridded, particle, and discrete data, logical groupings for discrete data, and an overall simplified and flexible interface for storing scientific data. MRAF is being used by the DOE at the Hanford-Meterological Site for observational data and will be used for their weather-modelling. SSESCO is also working with meterologists around the US and France who will be moving to netCDF before the end of the year (mostly users of savi3D).

The VISAGE visualization system, developed at the General Electric Corporate Research and Development, (Schroeder, WJ et al, "VISAGE: An Object-Oriented Scientific Visualization System", Proceedings of Visualization '92 Conference) uses netCDF as the preferred format. VISAGE is used at GE Corporate R&D, GE Aircraft Engine, GE Canada, GE Power Generation, as well as ETH Zurich, Switzerland, MQS In Chieti, Italy, and Rensselaer Polytechnic Institute in Troy, New York.

GE has another application called "Decimate" that does polygon reduction/decimation (Schroeder,WJ et al, "Decimation of Triangle Meshes", Proceedings of SIGGRAPH '92). Again, this application uses netCDF as a preferred format. Decimate is currently licensed to Cyberware, Inc., makers of 3D laser digitizing hardware. Decimate is currently bundled with the scanners, and will soon be available as a commercial product.

July 20, 1993

Contents

---

27. How do I make a bug report?

Q1: What is HDF?
A1: . HDF stands for Hierarchical Data Format. It is a multi-object file format for the transfer of graphical and numerical data between machines.
. HDF is a versatile file format. It supports six different data models. Each data model defines a specific type of data and provides a convenient interface for reading, writing, and organizing a unique set of data elements.
. HDF is a self-describing format, allowing an application to interpret the structure and contents of a file without any outside information.
. HDF is a flexible file format. With HDF, you can group sets of related objects together and then access them as a group or as individual objects. There are pre-defined sets for raster images and floating point multidimentional arrays. User can also create their own grouping structures using an HDF feature called vgroups.
. HDF is an extensible file format. It can easily accommodate new data models, regardless of whether they are added by the HDF development team or by HDF users.
. HDF is a portable file format. HDF files can be shared across platforms. An hdf file created on one computer, say a Cray supercomputer, can be read on another system, say IBM PC, without modification.
. HDF is available in the public domain.

Q2: What is the latest official release of HDF?
A2: HDF version 3.2 release 4 is the latest official release of HDF.

Q3: What does the latest official release of HDF contain?
A3: HDF3.2 release 4 contains the HDF library, HDF command-line utilities, and a test suite.

Q4: What is in the HDF library?
A4: HDF currently supports six data structure types: 8-bit raster images, 24-bit raster images, color palettes, scientific data sets(multi-dimensional arrays), text entries and vdatas (binary tables).
The HDF library contains a general purpose interface and six application level interfaces, one for each data structure type. These application level interfaces are specifically designed to read, write and manipulate one type.
The general purpose interface contains functions, such as file I/O, error handling, memory management and physical storage, etc. HDF library functions can be called from C or FORTRAN user application programs.

Q5: What are HDF command-line utilities?
A5: HDF command line utilities are application programs that can be executed by entering them at the command level, just like other UNIX commands.
They provide capabilities for doing things with HDF files for which you would normally have to write your own program. For example, the utility r8tohdf is a program that takes a raw raster image from a file and stores it in an HDF file in a raster image set.

Q6. How many platforms does HDF run on?
A6: HDF3.2 release 4 runs on the following platforms:

Platform	Machine Type
Sun-4	SUN

Cray-Y/MP	UNICOS
Convex	CONVEX
SGI Iris 4D	IRIS4
IBM RS/6000	IBM6000
DECstation 3100	MIPSEL
DECstation 5000	MIPSEL
Vax/VMS	VMS
HP 9000	HP9000
Macintosh	MAC
IBM PC/clone	PC
NeXT	NEXT
Convex	CONVEXNATIVE

Q7: Which NCSA tools can I use to view HDF objects?
A7: NCSA has also developed a suite of software tools for scientific visualization that are based on HDF. For the Mac, NCSA DataScope and NCSA Image can be used to view HDF files. On the PC, NCSA Audible Image uses HDF. X-based workstations can use XImage, NCSA XDataSlice and NCSA Polyview. NCSA Collage, a cross-platform collaborative tool, displays and manipulates HDF files. Version 2.0 of NCSA Mosiac, a networked information browser and World-Wide-Web client, can list and display the contents of HDF files.

Q8: Is there any commercial or public domain visualization software that accepts HDF files?
A8: Commercial software -- Spyglass, Wavefront, PCI, PV-Wave, IDL, AVS, Data Explorer and IRIS Explorer. Public domain software -- FREEFORM and GRASS and NCSA visualization tools.

Q9: What are the new features included in the next version of HDF?
A9: HDF 3.3 will add support for:

. reading and writing JPEG compressed 24-bit and 8-bit raster images
reading and writing hyperslabs from an SDS.
. reading and writing Little-Endian (PC Native) mode files on all platforms.
. integration of netCDF with HDF.
. a new multi-file SDS interface which supports "time-series" data as well as a more general metadata abstraction.

Q10: What is the status of the next version of HDF?
A10: HDF 3.3 Beta Release 1 is on NCSA anonymous ftp server. It has been tested on the following machines:

Platform	'base library'	'HDF / netCDF part'
Sun4	X	X
IBM/RS6000	X	X
SGI	X	X
Convex	X	X
Cray-Y	X	X
Cray-2	X	X
NeXT	X	X
HP	X	X
DecStation	X	X
DEC Alpha	X	
Fujitsu VP	X	
IBM PC - MSDOS	X	
IBM PC - Windows	X	

Q11: Are there any conversion programs available to convert non-HDF image files into HDF files or vise versa?
A11: HDF utilities r8tohdf, paltohdf, hdftor8, hdftopal, hdf24to8 and ristosds convert between raw raster images, palettes and hdf files.

X Window dump, Sun raster and 8-bit raw image files into HDF.
The SDSC Image Tools are tools developed at San Diego Supercomputer Center to handle image manipulation and file format conversion for a wide range of more than 20 file formats.

Q12: Where could I get source code and information relevant to HDF and HDF utilities?
A12: HDF and HDF utilities are public domain software. They are on the NCSA anonymous ftp server, in subdirectory HDF/.
The contents in HDF/ are:

```
README          describing files and subdirectories in ftp/HDF/.
FAQ             This file, frequently asked questions about HDF.
prev_releases   previous releases of HDF3.2r4, HDF3.1r5, Macintosh
                version of HDF3.1r3, HDF3.2r2, HDFVset2.1
HDF3.2r4/       HDF 3.2 release 4 (latest release)
HDF3.3-BETA/    HDF 3.3 Beta testing, release 1
contrib/        contributions from hdf users outside and inside NCSA
examples/       examples of hdf programs--good for testing, too
newsletters/    HDF newsletters
tarexamples/    compressed tar files of examples
HDFVset/        README  -- where to get old/new version of HDFVset
```

Q13: What are the files in HDF/HDF3.2r4 on ftp.ncsa.uiuc.edu?
A13: There are five subdirectories:

1. unpacked -- contains all source code.
2. patches -- bug fixes to the source.
   The patch files can be used to patch the source code in the subdirectory unpacked/.
3. tar -- contains compressed UNIX archived version of, the code in unpacked/. Users should check the patches/ directory to find out which files have been changed and decide whether or not the patches are needed for their applications.
4. hqx -- contains Macintosh BinHexed archived version of, the code in unpacked/. Users should check the patches/ directory to find out which files have been changed and decide whether or not the patches are needed for their applications.
5. zip -- contains IBM-PC compressed archived version of, the code in unpacked/. Users need to check the patches/ directory to find out which files have been changed and decide whether or not the patches are needed for their applications.
6. doc -- documentation for HDF and Vsets.

Each subdirectory has a README file explaining what files are in that directory and how to use those files.

Q14: How do I get source code for HDF library and utilities?
A14: You may obtain HDF via FTP, an archive server, or US mail.

FTP server:
If you are connected to Internet (NSFNET, ARPANET, MILNET, etc) you may download HDF source code at no charge from the anonymous ftp server at NCSA. The Internet address of the server is:

ftp.ncsa.uiuc.edu
or
141.142.20.50

Log in by entering anonymous for the name and enter your local e-mail address (login@host) for the password.

After logged in to ftp you need to change directory to HDF/HDF3.2r4/.
If you want packed source code, change directory to tar/ or hqx/ or zip/ (see A13 above). Files in those directories need to be transferred using binary mode.

If you want unpacked source code, change directory to unpacked/ and transfer all the files in unpacked/ and in its subdirectories to your host. Refer to HDF/HDF3.2r4/unpacked/README.FIRST for details.

If you have any questions regarding this procedure or whether you are connected to Internet, consult your local system administration or network expert.

Archive server:

E-mail a request to:
archive-server@ncsa.uiuc.edu
Include in the subject or message line the word "help", then press RETURN.
Send another e-mail request to:
archive-server@ncsa.uiuc.edu
Include in the subject or message line the word "index", then press RETURN.
The information you receive from both the help and index commands will give you further instructions on obtaining NCSA software. Refer to Chapter one of HDF Calling Interfaces and Utilities for details.

US mail:

A tape or CDROM archive of HDF is also available for purchase through the NCSA technical Resources Catalog. To obtain a catalog, contact:

NCSA Documentation Orders
152 Computing Applications Building
605 East Springfield Avenue
Champaign, IL 61820

(217) 244-4130

Q15: How do I install the HDF library on my computer?
A15: HDF 3.2 can be built with a single command from the top level directory where the subdirectories src/, util/ and test/ reside. The file Makefile.template is a generic, machine independent Makefile which you can modify if there is no Makefile already built for your machine.

For convenience, there are also machine customized makefiles. For example MAKE.IBM6000 is a Makefile suitable for compiling HDF on an IBM RS/6000. Assuming you are on an IBM RS/6000, copy MAKE.IBM6000 to Makefile and use the following commands to install different targets:

```
cp MAKE.IBM6000 Makefile

make
make allnofortran
 --- builds the HDF library and only the C interfaces, the
         utilities and the C test programs.
make all
```

utilities, and C and FORTRAN test programs.

Refer to the file INSTALL.ALL in the top level for making other targets.

Q16: How do I compile application programs that call HDF functions?
A16: To use HDF routines in your C program, you must have the line '#include "hdf.h"' near the beginning of your code. You may also need to include additional header files (eg. vg.h or dfsd.h, etc), depending on the interfaces you are using.

If you are on a SUN SPARC, the include files are in the directory "incdir", and the library file "libdf.a" is in "libdir". Use the following command to compile a C program "myprog.c":

```
cc -DSUN -Iincdir myprog.c libdir/libdf.a -o myprog
```
or
```
cc -DSUN -Iincdir myprog.c -o myprog -L libdir -ldf
```

HDF supports ANSI C. If you have ANSI C but ANSI C is not the default C compiler on your machine, the option -ansi should also be included in the command line shown above.

For FORTRAN programs, if your FORTRAN compiler accepts 'include' statements, you may include constant.i and dffunc.i in your program. Otherwise, you need to declair in your program all the constants used and functions called by the program.
To compile a FORTRAN program "myprogf.f" use:

```
f77 -o myprogf myprogf.f libdir/libdf.a
```

Q17: What documentation for HDF is available on the ftp server?
A17: Current HDF Documentation includes:
"HDF Calling Interfaces and Utilities" was designed for users who use HDF to store or manipulate their data. It describes the type of data each interface deals with and all the routines contained in each interface. HDF command-line utilities are also described in this manual. It has been updated to HDF3.2.

"HDF Specifications" was designed for those who need detailed information about HDF, such as HDF application program interface developers. It describes the basic structure, components, and software layers of HDF; specifies supported HDF tags and discusses the portability of HDF. It has been updated to HDF3.2.
The current version is a draft and is now undergoing editing. We expect a final version to be ready by mid-July.

We are working on the documentation for the HDF3.2 Vset interface. Currently, users may use "HDF Vset 2.0", "vset2.1.extra.doc" and the Vset section of HDF3.2 MigrationNotes as temporary reference for HDF Vset.

"HDF Vset 2.0" was designed for users who use HDF Vset to store and manipulate their data. It describes the organization of data within vdatas, vgroups and vsets, routines that manipulate vdatas and vgroups, and vset utilities that manipulate vsets.

"Vset2.1.extra.doc" describes routines added in Vset2.1 and not covered by "HDF Vset 2.0".

The "MigrationNotes" of HDF3.2 addresses differences between HDF3.2 and HDF3.1, and explains how to migrate existing application programs from HDF3.1 to HDF3.2.

following volumes:

Getting Started -- Overview of HDF with simple examples. Its draft is completed and is titled "Getting Started With NCSA HDF".

User's Guide -- Full coverage of all HDF routines and command-line utilities

Reference Manual -- An alphabetical listing of all HDF routines and command-line utilities

HDF Specifications -- Same as the current HDF Specifications

Q18: How can I obtain machine readable copies of the documentation?
A18: On the NCSA anonymous ftp server users can find the drafts for "HDF Specifications", "HDF Calling Interfaces and Utilities", and "Getting Started with NCSA HDF", as well as the documentation for HDF Vset. They are in different subdirectories within the directory Documentation/ (on ftp). There is a README file in each subdirectory, which explains briefly what is contained in that subdirectory and related information.

The following is a list showing where (which subdirectory) each documente resides in on NCSA ftp:

Documentation	subdirectory
HDF Specifications (Draft)	Documentation/HDF.Specs/
HDF Calling Interfaces and Utilities (Draft)	Documentation/HDF3.2/
HDF Vset 2.0	Documentation/HDF.Vset2.1/
vset2.1.extra.doc	Documentation/HDF.Vset2.1/
Getting Started with NCSA HDF (Draft)	Documentation/HDF_getting_started/
MigrationNotes	HDF/HDF3.2r4/unpacked/

"HDF Specifications", "HDF Calling Interfaces" and "HDF Vset 2.0" were written using MicroSoft Word. In each subdirectory the *.sit.hqx file is the MS Word version of the documentation and is stuffed using Stuffit.1.5. The *.asc.tar file is the ASCII text version of the MS Word file. It is missing all figures and formatting , so it is not very readable.

"Getting Started" was created using FrameMaker. This document is available only in postscript form.

MigrationNotes and vset2.1.extra.doc are ASCII text files.

Q19: How do I get hard copies of HDF documentation?
A19: NCSA accepts orders for hard copies of HDF Specs, HDF Calling Interfaces, HDF Vset 2.0 and Getting Started. Interested users should contact the office of NCSA Documentation Orders at:

(217)244-4130
docorder@ncsa.uiuc.edu (Internet)
docorder@ncsagate (BITNET)

or

NCSA
University of Illinois at Urbana-Champaign
605 Springfield Ave.
Champaign, IL 61820

Before the new documentation is officially published, NCSA supplies drafts of the documentation only. Don't be surprised if you see the big sign "DRAFT" on the top page when you receive the new documentation!

MigrationNotes is contained in HDF3.2 releases. Users having difficulties in getting these two files should contact softdev@ncsa.uiuc.edu.

Q20: Is there any documentation for HDF 3.3 Beta available?
A20: The README file in HDF/HDF3.3-Beta/ and README files in its subdirectories explain the differences between HDF3.2r4 and HDF3.3.Beta. Those README files and HDF3.2 documentation supply most information necessary to HDF3.3 users.

Q21: Can new versions of HDF read hdf files written using older version of the HDF library?
A21: Our goal is to make HDF backward compatible in the sense of that HDF files can always be read by new versions of HDF. We have succeeded in doing so up to HDF3.3 Beta, and will continue to follow the principle as much as possible. The table below lists the backward compatibility of HDF. Note that Vdata and Vgroup interfaces has been merged into since HDF3.2. Before then, they were a seperate library named Vset.

Interface	read HDF3.1	read HDF3.2	read HDF3.3
HDF3.1			
-RIS8	YES	YES	YES(except JPEG)
-RIS24	YES	YES	YES(except JPEG)
-PALETTE	YES	YES	YES
-ANNOTATION	YES	YES	YES
-SDS	Float 32 only	Float 32 only	Float 32 only
Vset 2.1			
-VData	YES	YES	YES
-Vgroup	YES	YES	YES
HDF3.2			
-RIS8	YES	YES	YES(except JPEG)
-RIS24	YES	YES	YES(except JPEG)
-PALETTE	YES	YES	YES
-ANNOTATION	YES	YES	YES
-SDS	YES	YES	YES(except general attributes)
-VData	YES	YES	YES
-Vgroup	YES	YES	YES
HDF3.3			
-RIS8	YES	YES	YES
-RIS24	YES	YES	YES
-PALETTE	YES	YES	YES
-ANNOTATION	YES	YES	YES
-SDS	YES	YES	YES
-VData	YES	YES	YES
-Vgroup	YES	YES	YES

(*) The HDF 3.3 interface supports two types of SDS objects. Previous libraries are able to read "old-style" SDS objects. See the HDF 3.3 Documentation for more information.

However, old HDF libraries are NOT always be able to read hdf files written by newer version HDF libraries. For example, HDF3.1 can not read 16-bit integer SDS's because HDF3.1 did not support this data type.

Q22: Can my application programs which work with old versions of the HDF library always be compiled with new versions of HDF?
A22: As HDF evolves some functions have to be changed or removed. For example, in HDF3.2 some functions' formal parameters which were passed by value in HDF3.1 have to be passed by reference in order to support new number types. When this happens, old application programs need to be modified so that they can work with the new library.

Our policy is as follows: Keep existing functions unchanged as much as possible; create new functions when necessary to accommodate new features; if a new function covers the feature of an existing old function, the old function should still be callable by old application programs; should an old function be phased out the users will be forewarned and encouraged to switch to the new function; an old function will be removed from the library only if it is in conflict with the implementation of new features.

Q23: Does HDF support data compression?
A23: HDF3.2 supports RLE (Run Length Encoding) and IMCOMP raster image compression schemes. In addition to RLE and IMCOMP, HDF3.3 supports JPEG compression for raster images. We plan to support compression for all number types in the future. However, no definite date has been set at this moment.

Q24: Is there a mailing list for HDF discussions and questions?
A24: hdfnews is a mailing list for HDF and its users to share ideas and information. HDF will announce new releases and events via newsletters through hdfnews. Users are welcome to make their comments, criticisms and suggestions on hdfnews as well.

With the help and support from the USENET coordinator and users, we have set up the news group sci.data.formats on the network. It has been an excellent forum for discussion of various data file formats.

Q25: How can a user share with other users HDF software that he or she has written, such as HDF utilities and ports to machines that NCSA does not support?
A25: Several users have ported HDF to their machines or developed their own utility programs to convert data between hdf and other file formats. They contributed their programs to us for HDF user community. Those programs are on NCSA ftp server in subdirectory:

HDF/contrib/

Q26: How do I contribute my software to HDF user community?
A26: Contact softdev@ncsa.uiuc.edu indicating that you would like to contribute your software to HDF user community. We will set up a directory for you to send the contribution package.

For other users' convenience, your contribution package should include the software itself, a Makefile if possible, a man-page, test programs and input data files for test. A README file is also required. It should describe briefly the purpose, function and limitation of the software; on which platforms and operating systems it runs; how to compile, install and test it; who and where to contact for commets, suggestions or bug reports.

Q27: How do I make a bug report?
A27: All bug reports should be sent to bugs@ncsa.uiuc.edu. Any comments,

Suggestions and questions should go to softdev@ncsa.uiuc.edu.

Attached below is a bug report template. It would be very helpful for us to locate and fix the bug if all the information inquired in the template are supplied by the reporter.

```
------------------ Template for bug report ------------------------
To: bugs@ncsa.uiuc.edu
Subject: [area]: [synopsis]   [replace with actual AREA and SYNOPSIS]

VERSION:
   HDF3.3 Beta 2, patch00

USER:
       [Name, phone number and address of person reporting the bug.
        (email address if possible)]

AREA:
       [Area of the HDF source tree affected, e.g., src, util, test,
         toplevel. If there are bugs in more than one AREA, please use
         a seperate bug report for each AREA. ]

SYNOPSIS:
       [Brief description of the problem and where it is located]

MACHINE / OPERATING SYSTEM:
       [e.g. Sparc/SunOS 4.1.3,  HP9000/730-HPUX9.01...]

COMPILER:
       [e.g. native cc, native ANSI cc, gcc 2.45, MPW, ...]

DESCRIPTION:
       [Detailed description of problem. ]

REPEAT BUG BY:
       [What you did to get the error; include test program or session
        transcript if at all possible.  If you include a program, make
        sure it depends only on libraries in the HDF distribution, not
        on any vendor or third-party libraries.  Please Be specific;
        if we can't reproduce it, we can't fix it. Tell us exactly what
        we should see when the program is run. ]

SAMPLE FIX:
       [If available, please send context diffs (diff -c)]

[PLEASE make your Subject(SYNOPSIS): line as descriptive as possible.]
[Remove all the explanatory text in brackets before mailing.]
[Send to bugs@ncsa.uiuc.edu or to:

       HDF, SDG at NCSA
       605 E. Springfield Ave.
       Champaign, IL 61820 ]

------------------ End of Bug Report Template --------------------
```

Software for Manipulating or Displaying NetCDF Data
23 June 1993

A number of freely-available and commercial software packages are available or under development for manipulating or graphically displaying netCDF data. Packages that are already available are listed first, followed by descriptions of software that is still under development.

MATLAB interface:

Researchers at the US Geological Survey in Woods Hole have developed an interface between Matlab and NetCDF called MEXCDF.

MATLAB is a software package that integrates numerical analysis, matrix computation, signal processing and graphical display.

MEXCDF is a interface between netCDF 2.0 and MATLAB 3.5 that invokes the complete C language netCDF interface, as described in the NetCDF Users Guide. This single MEX file allows MATLAB users to read, write, and manipulate netCDF data files in an efficient and compact manner without writing C or FORTRAN code.

In addition, the interface has been enhanced in several ways:

1. Dimensions and variables accessible by number or name.
2. Attributes accessible by number or name.
3. Parameters accessible by number or name.
4. Prepended "nc" not necessary for operation names.
5. Prepended "NC_" not necessary for specifying parameters.
6. Parameter names not case-sensitive.
7. Required lengths default to actual lengths via -1.
8. AutoScaling via "scale_factor" and "add_offset" attributes.

As an example, to read the following 2D array 'elev' in file 'foo.cdf',

```
short elev(lat, lon)
   elev:scale_factor=100.
   elev:add_offset=0.,
```

the required MATLAB commands using mexcdf are simply:

```
mexcdf('open','foo.cdf','nowrite');
elev=mexcdf('varget',cdfid,'elev',[0 0],[-1 -1]);
mexcdf('close');
```

The edges values "-1" means get all the values in this dimension, and scale_factor and add_offset are handled automagically by varget.

For more information regarding this software, get the file /pub/mexcdf/README via anonymous ftp from crusty.er.usgs.gov (128.128.19.19) or contact Rich Signell at rsignell@crusty.er.usgs.gov.

PolyPaint:

PolyPaint is an interactive 3D visualization package from NCAR. It has a rich suite of visualization tools that operate on three-dimensional netCDF data fields that follow certain PolyPaint-specific conventions. Three-dimensional iso-contours may be displayed as color-shaded surfaces. Features include multiple-object superposition, diffuse and spectral lighting from multiple light sources, transparency, haze, stereo viewing, and volumetric rendering, which may be combined with shaded surfaces. PolyPaint V 3.4 uses netCDF for data sets, and storage of geometry information. PolyPaint+ alpha-version, currently under development uses netCDF and DataHub from JPL. You can get more information on PolyPaint from boyd@mmm.ucar.edu. To get information on ordering a license for PolyPaint, contact Pat Waukau

(waukau@ncar.ucar.edu).

AVS:

An AVS module has been written that allows multi-dimensional netCDF data sets to read into AVS as uniform or rectilinear field files. The AVS user can point and click to specify the name of the variable in the selected netCDF file, as well as selecting the hyperslab. If 1D coordinate variables exist (a variable that has the same name as a dimension) then the coordinate variable will be used to specify the coordinates of resulting rectilinear field file. If no coordinate variable exists, then the resulting field file will be uniform. Once in AVS, there are hundreds of analysis and display modules available for image processing, isosurface rendering, arbitrary slicing, alpha blending, streamline and vorticity calculation, particle advection, etc. AVS runs on many different platforms (Stardent, DEC, Cray, Convex, E and S, SET, Sun, IBM, SGI, HP, FPS and WaveTracer), and it has a flexible data model capable of handling multidimensional data on non-Cartesian grids.

The module is available from the International AVS Center (avs.ncsc.org) in the directory /avs_modules/data_input/read_netcdf

Spyglass Dicer:

Spyglass Dicer is a commercial visualization tool for Apple Macintoshes that displays three-dimensional volumes of data in color, and that allows you to interactively create orthogonal slices, oblique slices, isosurfaces, blocks, and cutouts for viewing those data values. You can make all the data within a certain range of values invisible, create time or space animation sequences, print images in halftones or on color printers, and select from a variety of save and output features. Spyglass Dicer will read HDF and netCDF files as well as generic data. Contact Spyglass, Inc., P.O. Box 6338, Champaign, Il 61826. Phone: (217)355-6000, Fax: (217)355-8925.

IBM Data Explorer:

The IBM POWER Visualization Data Explorer, which runs on either an IBM RS/6000 or any of three models of the IBM POWER Visualization Server, can import netCDF data. The Data Explorer provides rendering of arbitrary slices from 3D data, color and opacity mapping of surfaces and volumes, isosurfaces, contours, streamlines, and various data manipulation functions. Contact Lloyd Treinish (lloydt@watson.ibm.com) for more information.

Ynot:

Unidata has contracted with MacDonald Dettwiler to develop Y0 (pronounced "Why not?"), a spatial data analysis and visualization system that applies the Model-View-Controller and spreadsheet paradigms to provide a "scientist's workbench" for data. Y0 can read and write netCDF files, and can display 2D data slices using color or dithered images and contouring. A beta-release is now available to licensed university sites. A later version will be generally available to all licensed educational Unidata sites, and MacDonald Dettwiler may market a commercial version as well. University sites that are interested should contact support@unidata.ucar.edu.

MGG_CDF

Lamont-Doherty Earth Observatory of Columbia University has converted all Marine Geophysics data (gravity, magnetics and bathymetry) acquired in the past 40 years by scientists at L-DEO as well as at other institutions to netCDF. A package of programs to access, maintain and display those files has also been completed. The software package is available via FTP (pub/cdf/MGG_CDF.tar.Z on lamont.ldeo.columbia.edu)

FREUD

Freud is an Open Look/X based interactive, object oriented framework for visualizing two dimensional data sets in netCDF form using NCAR graphics. Three objects, maps, contours, and vectors, are currently supported. Freud supports interactive manipulation of a number of parameters, including map display area, map projection, contour values, style, fill pattern, and color. Parameters can be saved for future reuse. Freud is hierarchically organized; defaults for most items are automatically chosen by the program, and panels for adjusting options are only displayed when the user explicitly requests them. The program operates on Sun or DEC Mips workstations. For more information, contact Joe Sirott (sirott@atmos.washington.edu).

Wavefront Data Visualizer:

Wavefront's Data Visualizer product provides isosurfaces, volume visualization, "smoke wands", 3D trajectories, animation and other visualization capabilities underneath a nice point and click interface. The product will accept netCDF data.

EPIC:

NOAA's Pacific Marine Environmental Laboratory (PMEL) has developed the EPIC software package for oceanographic data. EPIC provides graphical display and data field manipulation for multi-dimensional netCDF files (up to 4 dimensions). PMEL has been using this software on Unix and VMS for about a year. At present, they have:

* a data file I/O library (epslib, which is layered on top of the netCDF library).

* epslib allows transparent access to multiple data file formats

* a netCDF file calculator and data editor (see below)

* time series spectral analysis package (spectra, coherences, time and frequency domain eofs, rotary spectra, etc)

This software was developed on Sun/Unix and is also supported for DEC/Ultrix and VAX/VMS. It was developed as an expansion to the VAX/VMS based EPIC data management, display and analysis system for observational oceanographic time series and hydrographic data. The VAX EPIC software, about 100 programs for oceanographic display and analysis, is being ported to Unix as well, and they are interested in coordinating with others who may be developing oceanographic software for use with netCDF files. The EPIC software is available via anonymous FTP from csg.noaapmel.gov in the netcdf/ directory. Contact Nancy Soreide, nns@noaapmel.gov, for more information.

NetCDF File Calculator and Data Editor:

The netCDF calculator and data editor is part of the EPIC package, described above.

Algebraic manipulations available under the netCDF calculator include arithmetic operations (addition, subtraction, multiplication and division), exponentiation, log functions, square root, and some additional functions such as differentiation, integration, regridding and removing the mean value. In addition to the built-in functions, the calculator also allows user written routines for manipulation of data. The netCDF calculator also provides the ability to extract a subset of the EPIC System Data sets by specifying limits in any or all of the four dimensions (x,y,z,time) or (i,j,k,l).

The netCDF calculator includes an interactive data-editing function for one dimensional data sets. This displays the data and provides for replacement of individual points or a range of data points by linear interpolation or by typing in a replacement value. Data sets generated by the calculator data manipulation or data editing functions can be plotted with the standard PPLUS plotting commands and can also be written out as data files in any of the EPIC System formats (including netCDF).

The netCDF calculator uses an expression language, much like C: although there are several control-flow statements, most statements such as assignments are expressions whose value is disregarded. For example, the assignment operator "=" assigns the value of its right operand to its left operand, and yields the value, so multiple assignments work. The calculator knows about four different data types: scalar (double precision floating point), string, slab (a specification for a four dimensional hyper-slab), and fields (a sub-sampled region of a four dimensional data set). The syntax for each of the four types are similar, however, not all operations are valid for all data types.

The netCDF calculator discussed here is a simple programmable interpreter for floating point, string and field expressions. It has been extensively modified from hoc (The UNIX Programming Environment, Kernighan and Pike, Chapter 8). It has C-style control flow, function definition and the usual numerical built-in functions. The netCDF calculator has been developed using lex, a lexical analyzer, and yacc, a parser generator. This allows a systematic and consistent syntax to be implemented easily.

GMT:

The Generic Mapping Tools (GMT) were recently described in the article: "Free Software Helps Map and Display Data," by Paul Wessel and Walter H. F. Smith, EOS Transactions, American Geophysical Union, Volume 72, Number 41, October 8, 1991, page 441.

This article describes a package of tools and a library that can be used to manipulate columns of tabular data, time series, and gridded data sets and to display these data in a variety of forms ranging from simple x-y plots to maps and color, perspective, and shaded-relief illustrations. The package uses netCDF for representing two dimensional data on regular grids. GMT uses the PostScript page description language which can create arbitrarily complex images in gray tones or 24-bit true color by superimposing multiple plot files. Line drawings, bitmapped images, and text can easily be combined in one illustration. GMT software is written as a set of UNIX tools and is totally self-contained [except for needing netCDF] and fully documented. The system is offered free of charge to federal agencies and nonprofit educational organizations worldwide. It's available from kiawe.soest.hawaii.edu in pub/gmt/gmtv2.0.tar.Z

WXP:

The Purdue Weather Processor. WXP is an integrated system for decoding, analyzing, and displaying various types of meteorological information captured by Unidata's Local Data Manager system. WXP decodes broadcast weather data and model outputs into WXP-specific netCDF files, and provides analysis and display application programs for manipulating and displaying the data in those files. Universities must be licensed by Unidata and Purdue to obtain the WXP software. For information on obtaining a WXP license, contact support@unidata.ucar.edu.

HDF Interface:

A group at the The National Center for Supercomputing Applications (NCSA) has added the netCDF interface to their Hierarchical Data Format (HDF) software. HDF is an extensible data format for self-describing files. A substantial set of applications and utilities based on HDF is available; these support raster-image manipulation and display and browsing through multidimensional scientific data. An implementation is now available that provides the netCDF interface to HDF. With this software, it is possible to use the netCDF calling interface to place data into an HDF file. The netCDF calling interface has not changed and netCDF files stored in XDR format are readable, so existing programs and data will still be useable (although programs will need to be relinked to the new library). There is currently no support for the mixing of HDF and netCDF structures. For example, a raster image can exist in the same file as a netCDF object, but you have to use the Raster Image interface to read the image and the netCDF interface to read the netCDF object. The other HDF interfaces are currently being modified to allow multi-file access, closer integration with the netCDF interface will probably be delayed until the end of that project.

Eventually, it will be possible to integrate netCDF objects with the rest of the HDF tool suite. Such an integration will then allow tools written for netCDF and tools written for HDF to both interact intelligently with the new data files. For more information, contact Chris Houck (chouck@xongmao.ncsa.uiuc.edu).

SIEVE:

Researchers at the USGS in Reston, VA are developing the Scientific, Interactive and Extensible Visualization Environment -- SIEVE, which allows the user to select interactively multi-dimensional arrays of scalars or of vectors from netCDF files and to plot them using NCAR Graphics and XGKS. Plots may be displayed in an X Window or saved to a CGM. An abstract describing the current state of SIEVE is available via anonymous ftp from sparky.er.usgs.gov in the pub directory as sieve.ps. Contact hjenter@stress.er.usgs.gov for more information on this project.

NetCDF Operators:

The Unidata Program Center is implementing a suite of netCDF programs and utilities for manipulating netCDF data in UNIX-like environments. A draft functional and syntactical specification is available via anonymous FTP from unidata.ucar.edu in pub/sdm/ncprogs.ps (PostScript) or pub/ sdm/ncprogs.txt (ASCII). The initial release, now available, consists of three netCDF operators constituting the "proof of concept" operator set, an associated support library, and documentation. The netCDF operators are intended to:

* Support the sharing of data by acting as front- and back-end filters to other, more comprehensive data analysis packages;

* Perform quick and dirty data analysis and manipulation; and

* Provide examples of generic netCDF programs.

FERRET:

An ocean modelling group at NOAA's PMEL will soon make available a visualization and analysis package called FERRET for large, gridded data sets. FERRET supports generic netCDF files over a fairly broad range. FERRET provides a collection of basic transformations (averaging, integrating, smoothing, arithmetic and trigonometric operators, etc.) and a familiar, mathematics-like language for interactively defining new variables from older ones. FERRET's graphical and analytical tools can be applied as readily to variables defined by the user as they can to "raw" data. FERRET's graphical tools include animations, color shaded plots, contour plots, line and scatter plots, 3-dimensional fish nets, multiple windows and overlays. FERRET treats all four axes of the space-time coordinate system as equivalent. No special commands are required and no limitations are imposed when working with time series data. FERRET automatically provides full labelling and titling of all plots with information pulled from the data in a self-describing format. FERRET has built-in memory and disk management capabilities to handle data sets too large to fit within the limitations of on-line storage. FERRET is written in standard FORTRAN 77 with graphics based on the ISO GKS standard. It is currently running on VAX/VMS and DEC Ultrix computers with a port to Sun UNIX underway. Contact Steve Hankin, hankin@noaapmel.gov, for more information.

Envision:

Envision is an interactive system for the management and visualization of large scientific data sets. It currently runs on UNIX workstations under X/Motif, manages data stored in netCDF files, and does visualization using NCSA XImage, NCSA Collage, and IDL.

Envision is public domain software and is now available by anonymous FTP. The primary ftp site for Envision is vista.atmos.uiuc.edu in pub/envision. A secondary FTP site is csrp.tamu.edu. Included in this release are binaries for IBM RS6000 (on which it is developed), HP, Sun, and SGI, as well as source and make files. Complete documentation, along with samples of data and Envision projects are also provided. Envision is software currently undergoing development and this is the first beta test release. Help is requested in reporting any bugs to envision@vista.atmos.uiuc.edu. Suggestions for Envision's improvement are also welcome.

Envision is being jointly developed at the Department of Atmospheric Sciences at University of Illinois at Urbana-Champaign and the Department of Meteorology at Texas A&M University. Funding for this project is from NASA's Applied Information Systems Research Program.

For further information, contact Keith Searight (keith@vista.atmos.uiuc.edu).

IDL interface:

IDL (Interactive Data Language) now supports data in netCDF format. IDL is a commercial product for the interactive analysis and visualization of scientific and engineering data available from

Research Systems, Inc.
777 29th Street, Suite 302
Boulder, CO 80303
(303) 786-9900

As an example, here is how to read data from a netCDF variable named GP in a file named "data/aprin.nc" into an IDL variable named gp using the IDL language:

```
id = ncdf_open('data/april.nc')
ncdf_varget,id, ncdf_varid( id, 'GP'), gp
```

Now you can visualize the data in the gp variable in a large variety of ways and use it in other computations in IDL. You can FTP a demo version of IDL, including the netCDF interface, by folowing the instructions in pub/idl/README available via anonymous FTP from gateway.rsinc.com or boulder.colorado.edu.

NCAR Graphics:

The Scientific Visualization Group of the NCAR Scientific Computing Division is developing the NCAR Command Language (NCL) for interactive data manipulation and plot specification. This language is being developed in support of a project to create, from existing utilities, a scientific visualization environment that will provide a means of reading and writing data, a means of manipulating that data, and a means of visually analyzing the data interactively.

NCL is intended to provide easy and intuitive access to datasets and allow users to explore and process their data prior to visualization. Since datasets often come in a variety of data formats, grid sizes, grid resolutions, and units, very different datasets often need to be combined, compared, and used at the same time. Currently, specialized applications must be developed to read individual datasets and transform them into a form that is compatible with other datasets being used, as well as with the graphics package being used.

NCL allows different datasets, in different storage formats, to be imported into one uniform and consistent data manipulation environment. The primary data format used internally by NCL is the netCDF data format. NCL doesn't place any restrictions or conventions on the organization of input netCDF files.

NCL is a complete programming language that provides flexibility and configurability. In NCL the primary data type is the data file record. A data file record stores one or more variables, dimension information, coordinate variable information and attribute information as one NCL object. A binary file can be read in, dimension names, variable names, attributes and coordinate variables can be assigned to it using NCL language constructs and the resulting file record can be written to any of the currently supported formats, including netCDF, without writing a single line of source code.

The NCL language also provides the ability to specify configuration parameters for visualizations of data. These descriptions can be stored and used as user defaults and reused as plot specifications.

The function set of NCL contains built-in data processing and mathematical functions and can be extended by the user to provide custom data processing techniques, as well as custom data ingestion.

Contact Ethan Alpert, the NCAR Interactive project coordinator, at ethan@ncar.ucar.edu for more information. The beta test version of NCL should be available by 1/93.

February 1991, National Center for Atmospheric Research, Boulder, Colorado, 80307-3000]

Advantages and disadvantages of HDF and netCDF

by Glenn Davis and Russ Rew

Both the Hierarchical Data Format (HDF) and Network Common Data Form (netCDF) software packages are designed to make sharing data among different platforms easier. Both are now available on the CRAY Y-MP8/864. (See "HDF Version 10.3 installed on shavano" and "NetCDF Version 1.10 installed on shavano" in this issue for more information.) This article introduces the two packages and discusses their most significant differences.

HDF

The National Center for Supercomputing Applications (NCSA) has developed the HDF software and made it freely available. HDF is an extensible data format for self-describing files. A substantial set of applications and utilities based on HDF is available; these support raster-image manipulation and display and browsing through multidimensional scientific data. The HDF software includes a package of routines for accessing each HDF data type, as well as a lower-level interface for building packages to support new types. HDF supports both C and Fortran interfaces, and it has been successfully ported to a wide variety of machine architectures and operating systems. HDF emphasizes a single common format for data, on which many interfaces can be built.

NetCDF

Unidata has developed the netCDF software and made it freely available. NetCDF allows you to create, access, and share data in a form that is both self-describing and machine-independent. Unidata software uses netCDF as an interface between system-level programs that capture broadcast meteorological data and application programs that analyze and display it. Other groups have found netCDF useful for sharing data among different architectures, providing a flexible way to access data cross-sections, or providing an extensible interface to data that insulates application programs from details of data formats. Like HDF, netCDF supports both C and Fortran interfaces, and it has been successfully ported to a wide variety of machine architectures and operating systems. NetCDF emphasizes a single common interface to data, implemented on top of an architecture-independent representation.

Machine independence

Usually the data in HDF files are stored in a machine-independent format. You can override this with an HDF library call that instructs HDF to store the data in the machine's native format. This option is currently only available for Cray's 64-bit native format.

HDF currently supports two types of multidimensional arrays--Cray floating point and IEEE 32-bit floating point--whereas netCDF offers a much fuller range of multidimensional arrays. NetCDF multidimensional arrays correspond to the native integer and floating-point types available on the platforms on which it is implemented. The associated machine-independent data types are characters, eight-bit bytes, 16-bit integers, 32-bit integers, 32-bit IEEE floating point, and 64-bit IEEE floating point.

NetCDF data are machine-independent: the form in which characters, integers and floating-point data are stored is the same for netCDF data written on any platform on which the netCDF library is implemented. This has both advantages and disadvantages when compared with HDF. Machine-independent data can be shared transparently across network file systems. Conversion to native form only occurs on the data as they are accessed, so converting large files of data from one form to another before accessing a small subset of the data is an unnecessary step. On the other hand, the conversion to and from the machine-independent external data representation happens on every data access, so I/O may be slower and unnecessary conversions may occur when netCDF files are used on only one type of platform. HDF is significantly faster than netCDF in accessing large arrays of floating point numbers on a Cray computer (or on other architectures that do not use IEEE-standard floating-point numbers), but the performance difference for other types of data is less significant.

Data tags

HDF uses a centralized registry of basic data tags. The HDF tags designate fundamental data types needed in application programs.

NetCDF data has no such centralized registry of data tags. The structure of netCDF data is stored with the data. This has both advantages and disadvantages when compared with HDF. NetCDF can be used to store data in unanticipated forms, without registering the structure or variable names first, so it is more flexible for some purposes. HDF can identify complex data structures with a small tag, so it can store such structures more compactly than netCDF. Generic programs that deal with many different kinds of data are easier to write if all the structures they will deal with are known in advance, as is the case for HDF data tags. Programs that deal with data that carries its own description, such as netCDF, may adapt more easily to changes.

Interface size

HDF is a single format that supports a number of interfaces for raster images, color palettes, annotation, scientific datasets, general-purpose data access, and utilities. There are more than 75 routines in the HDF C interface. Many application-level programs are available that use the HDF format, especially for scientific visualization.

In contrast, netCDF is a single interface for many different kinds of data. One of the aims of the netCDF interface is to have a small ''surface area,'' so that the interface is easy to learn and use. There are only 28 routines in the netCDF C interface. There are currently only a few general-purpose netCDF utilities, although commercial scientific visualization applications that accept netCDF data are now available.

The future

The NCSA group that developed and maintains HDF software has announced their intent to support the netCDF interface for scientific data access, if it is practical to implement it on top of the basic HDF layer. Such an implementation would eventually permit applications software written using either the netCDF or HDF scientific dataset interfaces to be shared among both communities of users.

[Glenn Davis is a systems programmer and Russ Rew is the head of the Systems Group at Unidata in Boulder, Colorado. Thanks to Mike Folk at NCSA for reviewing this article.]

[This is an ASCII version of an article that appeared in SCD Computing News, February 1991, National Center for Atmospheric Research, Boulder, Colorado, 80307-3000]

Advantages and disadvantages of HDF and netCDF

by Glenn Davis and Russ Rew

Both the Hierarchical Data Format (HDF) and Network Common Data Form (netCDF) software packages are designed to make sharing data among different platforms easier. Both are now available on the CRAY Y-MP8/864. (See "HDF Version 10.3 installed on shavano" and "NetCDF Version 1.10 installed on shavano" in this issue for more information.) This article introduces the two packages and discusses their most significant differences.

HDF

The National Center for Supercomputing Applications (NCSA) has developed the HDF software and made it freely available. HDF is an extensible data format for self-describing files. A substantial set of applications and utilities based on HDF is available; these support raster-image manipulation and display and browsing through multidimensional scientific data. The HDF software includes a package of routines for accessing each HDF data type, as well as a lower-level interface for building packages to support new types. HDF supports both C and Fortran interfaces, and it has been successfully ported to a wide variety of machine architectures and operating systems. HDF emphasizes a single common format for data, on which many interfaces can be built.

NetCDF

Unidata has developed the netCDF software and made it freely available. NetCDF allows you to create, access, and share data in a form that is both self-describing and machine-independent. Unidata software uses netCDF as an interface between system-level programs that capture broadcast meteorological data and application programs that analyze and display it. Other groups have found netCDF useful for sharing data among different architectures, providing a flexible way to access data cross-sections, or providing an extensible interface to data that insulates application programs from details of data formats. Like HDF, netCDF supports both C and Fortran interfaces, and it has been successfully ported to a wide variety of machine architectures and operating systems. NetCDF emphasizes a single common interface to data, implemented on top of an architecture-independent representation.

Machine independence

Usually the data in HDF files are stored in a machine-independent format. You can override this with an HDF library call that instructs HDF to store the data in the machine's native format. This option is currently only available for Cray's 64-bit native format.

HDF currently supports two types of multidimensional arrays--Cray floating point and IEEE 32-bit floating point--whereas netCDF offers a much fuller range of multidimensional arrays. NetCDF multidimensional arrays correspond to the native integer and floating-point types available on the platforms on which it is implemented. The associated machine-independent data types are characters, eight-bit bytes, 16-bit integers, 32-bit integers, 32-bit IEEE floating point, and 64-bit IEEE floating point.

NetCDF data are machine-independent: the form in which characters, integers and floating-point data are stored is the same for netCDF data written on any platform on which the netCDF library is implemented. This has both advantages and disadvantages when compared with HDF. Machine-independent data can be shared transparently across network file systems. Conversion to native form only occurs on the data as they are accessed, so converting large files of data from one form to another before accessing a small subset of the data is an unnecessary step. On the other hand, the conversion to and from the machine-independent external data representation happens on every data access, so I/O may be slower and unnecessary conversions may occur when netCDF files are used on only one type of platform. HDF is significantly faster than netCDF in accessing large arrays of floating point numbers on a Cray computer (or on other architectures that do not use IEEE-standard floating-point numbers), but the performance difference for other types of data is less significant.

Data tags

HDF uses a centralized registry of basic data tags. The HDF tags designate fundamental data types needed in application programs.

NetCDF data has no such centralized registry of data tags. The structure of netCDF data is stored with the data. This has both advantages and disadvantages when compared with HDF. NetCDF can be used to store data in unanticipated forms, without registering the structure or variable names first, so it is more flexible for some purposes. HDF can identify complex data structures with a small tag, so it can store such structures more compactly than netCDF. Generic programs that deal with many different kinds of data are easier to write if all the structures they will deal with are known in advance, as is the case for HDF data tags. Programs that deal with data that carries its own description, such as netCDF, may adapt more easily to changes.

Interface size

HDF is a single format that supports a number of interfaces for raster images, color palettes, annotation, scientific datasets, general-purpose data access, and utilities. There are more than 75 routines in the HDF C interface. Many application-level programs are available that use the HDF format, especially for scientific visualization.

In contrast, netCDF is a single interface for many different kinds of data. One of the aims of the netCDF interface is to have a small ''surface area,'' so that the interface is easy to learn and use. There are only 28 routines in the netCDF C interface. There are currently only a few general-purpose netCDF utilities, although commercial scientific visualization applications that accept netCDF data are now available.

The future

The NCSA group that developed and maintains HDF software has announced their intent to support the netCDF interface for scientific data access, if it is practical to implement it on top of the basic HDF layer. Such an implementation would eventually permit applications software written using either the netCDF or HDF scientific dataset interfaces to be shared among both communities of users.

[Glenn Davis is a systems programmer and Russ Rew is the head of the Systems Group at Unidata in Boulder, Colorado. Thanks to Mike Folk at NCSA for reviewing this article.]

OCEANOGRAPHIC CONVENTIONS for NETCDF
Draft no.1 - 10/7/92

Phase 1 - identifying issues

CONTENTS:

Introduction.

This file, cdfprofile.oceanography, will document a profile of conventions to standardize the usage of netCDF for many oceanographic data applications. The primary goal of this standardization is to facilitate data interchange. The document is expected to develop in three phases:

Phase 1 - We will enumerate the issues relevant to oceanographic data storage with netCDF and provide several alternative resolutions to issues. In Phase 1 **all** aspects of this document are open to comments and revision.

Phase 2 - Through email dialog we will discuss the resolutions of the issues that have been identified working towards a consensus on each issue. In Phase 2 the basic layout and strategy of the document will be fixed, however all technical content will be revisable. New issues will be added only if there is general agreement that they are vital to the document. (At this stage the dialog may be shifted off of netcdfgroup.)

Phase 3 - We will edit the document removing ambiguities and producing stable, readable text. Issues that become apparent during implementations must also be resolved at this time.

The entire process of arriving at a standardized profile should be open and should reflect the views of all members of the "oceanographic community" (an admittedly ambiguous term) who wish to participate. If a consensus cannot be reached on some issues it may be necessary to formalize a "voting" procedure to resolve the issues. (These procedures could be defined in Phase 1 of the document? volunteer?)

The scope of this work is broader than what we can hope fully to achieve. Some issues may need to be classified as "Beyond the scope of this document". Again, we should try to reach general agreement before classifying an issue this way.

Producing a final document that unambiguously describes all of the issues and resolutions will clearly be a significant piece of work. This can only be accomplished if we each offer complete and concise text when we make a contribution. I (Steve Hankin) will volunteer to serve as the document editor - pulling our contributions together into a single document and making it available via email and/or anonymous ftp. Since this is to be an entirely open process please speak your mind if you know of a preferable prospect for document editor.

This document presumes the contents of "conventions.info" (unidata.ucar.edu anonymous ftp directory pub/netcdf) and will not duplicate what is already described there. As both conventions.info and profile.oceanography will be evolving in parallel we will need to coordinate the documents throughout their evolutions.

Guidelines for Creating Profiles.

In the process of discussing issues and comparing alternative resolutions an explicit set of "guiding principles" would be an asset. Such principles include (please extend):

- o keep it simple (avoid proliferation of attributes)
- o minimize restrictions (don't reduce functionality)
- o profile-compliant files should remain intelligible to applications that know nothing of the profile (where possible)

Guidelines for Robust Applications.

Application programs will in general be far more restrictive in scope than the conventions described herein. These application programs can still perform useful work on many netCDF files that observe the conventions if they observe the following "motherhood" rules:

- o meta rule: don't crash, don't give up if possible
- o Uninterpretable attributes should be ignored
- o Variables with unsupported data types should be ignored
- o Applications should not assume particular units will be attached to particular variable names.
- o Applications that require recognized variable names should ignore variable names they do not recognize
- o Applications should avoid assumptions about the structure of the netCDF file:
  - dimensions may be defined which are unused
  - variables may use dimensions which have no corresponding coordinates defined
  - etc. (expand list)

Oceanographic Profile Issues.

1) Time axis representation

The file "conventions.info" suggests (e.g.)

```
variables:
    double time(nobs);
     time:units = "milliseconds since (1992-9-16
                     10:09:55.3 -600)"
```

(This will be implemented shortly in the udunits library.) Should we impose restrictions on data types (double, float, etc)? How should we standardize the format for the date string? (is this specificed by udunits?)

2) How to determine the orientation of a coordinate variable

The orientation of coordinate axes can be specified through a variety of mechanisms: agreed-upon names such as "lat","lon", etc.; implicit orientations inferred from the ordering of dimension names within a variable definition; orientations inferred from the units of the coordinate variable. None of these mechanisms appear to be adequate

Alternative 1:
Minimal restrictions on the naming of coordinate variables and choice of units. Applications should apply a multi-step algorithm to identify orientation as follows:
First - check the units of the coordinate variable:
Do the units imply a unique orientation (e.g. units of time, "degrees longitude", "layer", etc.) ?
If no, then check the name of the coordinate variable:
Does the variable name match a template (e.g. *depth*, *lon*, *lat*, *time*, x*, y*, z*, t*, etc.)?

Is this approach too complex? What about cases where the orientation remains ambiguous?

Alternative 2:
Introduce a variable attribute 'orientation' with a suitable naming convention for orientation strings (e.g. "west-east", "south-north")

Should this be an optional attribute that can be applied when the Alternative 1 technique fails?

3) Indicating Missing Data
Two attributes for missing data have been suggested: missing_value and _FillValue. The missing_value attribute has been dropped in netCDF version 2.0. Is there a need to support both attributes?

4) Case-insensitive Names
Should application programs be case-sensitive with respect to attribute and variable names? Should variable and attribute names within a single file be required to be case-insensitive-unique? (This refers to the **names** only; the values of string attibutes such as units would remain case-sensitive.)

Alternative 1: Case-insensitive. The peculiarities of Unix and C, while familiar to programmers, are not necessarily comfortable for users. Publication and conversation are complexified by case-sensitive names.

Alternative 2: Case-sensitive. Case-insensitivity would lead to incompatibilities with non-oceanographic netCDF files. There are conveniences to the use of e.g. "time", "Time", and "TIME" within the same file.

5) Multiple Time Axes in a File
Is there a need for multiple time axes defined within a single netCDF file? Or is there a reason to limit files to a single time axis? (Multiple time axes would conflict with some time encodings that have been discussed that involve global variables.)

Alternative 1: Permit multiple time axes (no conflict with time axes as suggested in conventions.info).

6) Need a global attribute to indicate profile type and revision
There should be a global attribute informing application programs explicitly what netCDF profile and revision a file adheres to. This issue needs to be addressed at a level higher than this oceanographic profile but some recommendations would be appropriate.

Alternative 1:

```
:profile = "oceanography";
:profile_version = 1.0;
```

7) Standardized (Conventional) Variable Names
The meteorological community has suggested a list of standardized variable names (see conventions.info). Should this list be extended to include additional oceanographic variables? How should these names fit this into the framework of "resources" as described in conventions.info? (We need input from folks familiar with "resources" in this context.)

8) Name String Lengths
Should attribute and variable names be further restricted with respect to length beyond the limit of 'MAX_NC_NAME' described in conventions.info?

Alternative 1: a practical limit of (say) 32 characters should be imposed. This is consistent with most programming languages. It simplifies the formatting burdens on applications. It does not prevent application programs from supporting longer names.

Alternative 2: Any limit other than the default limit of MAX_NC_NAME (128) could lead to incompatibilities with non-oceanographic netCDF files.

9) Multiple coordinate variables of same orientation
Is there a need to support multiple coordinate variables of the same orientation in a single netCDF file? (such multiplicity would preclude the use of strict names such as "lat" to designate geographical coordinate variables though templates like *lat* would still be possible)

Alternative 1: yes, there is a need (e.g. multiple current meter arrays with differing deployment depths; in modelling it is often desirable to compare results computed on numerous different axes of the same orientation - restrictions on naming of axes could be very inconvenient)

10) Requiring non-coordinate variables to be 4 dimensional
Is it acceptable to insist that all non-coordinate variables be represented as 4-dimensional (lat/long/depth/time) structures? Should there be other restrictions on number of axes?

Alternative 1: dimensionality should not be restricted to exactly 4 - the restriction would preclude some data types and would force misrepresentation of others. Some restriction on the maximum number of dimensions for a variable would, however, ease the burden on application writing.

11) Mandatory ordering of geographical dimensions
Is it acceptable to mandate that if dimensions with geographical significance are used in defining a variable they will be ordered as lat-lon-depth-time (i.e. time as the slowest moving axis)?

Alternative 1: yes with reservations - are there serious performance penalties?

Alternative 2: no - applications require greater flexibility than this. Perhaps a standard ordering could be defined and an attribute introduced that would indicate permutations. Example:

```
var:permutation = "TXYZ";
```

12) Coordinate Systems
As mentioned in conventions.info there is work underway at unidata on this subject leading towards the development, presumably, of a collection of conventional attributes and a new Unidata library, 'udgeoref'. Is this work sufficient for oceanographic data? Is this beyond the (initial) scope of this document?

13) Application-specific attributes
Would it be useful to standardize a collection of attributes that would coach application programs in areas not directly related to the data content - for example attributes that recommended display techniques such as

```
preferred_display_style="contour"
preferred_display_map="spherical polar"
```

Candidates? ...

14) Climatological Axes
What is the best method to represent a climatological time axis?

Alternative 1: attach the (boolean) variable attribute "periodic" to the time coordinate axis indicating the axis ends "join" modulo-fashion (this solution is useful for any periodic axis - also applicable to longitude). What about the base-date string (see issue 1)?

```
time: periodic = " ";
```

Alternative 2: Like alternative 1 but the attribute should indicate the "branch points" of the periodicity:

```
time: periodic_values = 0.,365.;
```

15) Use of Boolean Attributes
Issue 14 raises the general question of the appropriateness of boolean attributes (whose presence or absence indicates a modal state). There is no explicit mechanism in netCDF for creating a value-less attribute (see Issue 14 Alternative 1). Should profile.oceanography avoid boolean attributes? Or is this largely an aesthetic issue of the appearance of CDL files? Could CDL be extended in a future revision to support e.g.

```
time: periodic;
```

16) Vertical axis orientation
Often oceanographic data is organized with positive down on vertical axes. What is the best mechanism to indicate this in a netCDF file? (A similar question arises on latitude axes which may be south-positive or north-positive.)

Alternative 1: Introduce a (boolean) coordinate variable attribute "reversed".

Alternative 2: Combine this property together with others that have been discussed in a new attribute

```
depth: properties = "reversed, coordinates, vertical";
```

17) Longitude axis encodings

Longitudes encodings are not ... continuous across the dateline or continuous across the prime meridian; either westward or eastward may be positive; the range may be -180 to 180 or 0 to 360 or some other choice. How should netCDF convey this encoding?

Alternative 1: 4 variable attributes applied to the longitude coordinate variable:
- "reversed" for X positive, westward
- "discontinuity"=value (always give the minimum value)
- one of "Greenwich=value" or "dateline=value"

e.g. To define a longitude axis from 0 to 360, positive eastward, with zero representing Greenwich

```
       variables:
          float lon(lon);
           lon:Greenwich=0.;
           lon:discontinuity=0.;
```

Alternative 2: modify Alternative 1 by replacing the "discontinuity" attribute with

```
           lon:periodic_values = 0., 360.;
```

18) Unequally spaced coordinates
Is the location of grid points sufficient information to fully describe a coordinate axis with irregularly-spaced points? Or do we need auxiliary machinery to represent the boundaries between points?

Alternative 1: There are cases that require explicit boundaries between cells on an axis e.g. data collected in unequal bins. Is this a special-purpose need beyond the scope of this document?

19) Huge Data Sets / Multiple Files
Should we provide a standardized mechanism for associating multiple files in a single "project"? How should it function? as a time axis distributed among files? as multiple variables distributed among files? Is this beyond the scope of this document?

Alternative 1: a "parent" netCDF file with variables and attributes suitably defined to point to "child" files.

Alternative 2: a file naming convention such as

```
     my_cdf.001, my_cdf.002, my_cdf.003, ...
```

that will implicitly concatenate netCDF files along their record (or time?) axis.

20) Representing Sigma Coordinate Systems
How should variables defined on sigma coordinate grids be represented? Is this question within the scope of this document? Will it covered by the 'udgeoref' library?

Alternative 1:
A variable defined on a sigma coordinate system should possess an attribute "sigma". The coordinate variable corresponding to the vertical dimension should exist and have simple enumerated values 1, 2, ..., n. The coordinate variable should further have an attribute "sigma_positions" (?better name?) which gives the name of a variable containing the z coordinates. The z coordinate variable should be defined on the same dimensions as the original variable. e.g.

variables:

```
float    u(lat,lon,level,time); //
         u: sigma = " ";
integer level(level);
         level: sigma_positions = "depths";
float    depths(lat,lon,level); // time may be a
                                    dimension, too//
         depths:units="meters";
```

****

Real-Time and Shipboard data collection?
What are the special issues?
How to represent a cruise track? (** a requirement? **)
How to store variables with differing sampling intervals?
(Beyond Scope?)

Arctic oceanography
What are the special issues?

Climate research
What are the special issues?

Chemical oceanography
What are the special issues?

Biological Oceanography
What are the special issues?

Compressed data
Are there special cases where compression can fit a general framework?

Other special topic issues?

Steve Hankin | NOAA/PMEL | ph. (206) 526-6080
| 7600 Sand Point Way NE | FAX (206) 526-6744
| Seattle, WA 98115-0070 | hankin@noaapmel.gov

Date: Thu, 26 Aug 93 24:14:00 EDT
From: R.FINDLEY <R.FINDLEY@telemail>
Message-Id: <DGJD-5865-4170@telemail>
Subject: Do we need data distribution standards?
To: <findley>
Sender: Telemail-Daemon

Posted: Thu, Aug 26, 1993 12:14 PM EDT Msg: DGJD-5865-4170
From: R.FINDLEY
To: unols.rvtec
Subj: Do we need data distribution standards?

The problem from where I sit.

Data is being collected onboard RV Columbus Iselin at a rate of about 30 Mbyte per day. (Where does all this come from? 3 Mbyte from 150 kHz ADCP, same for 600 kHz ADCP, 1 - 3 Mbyte from CTD system, 1 - 2 Mbyte per day from CIDS (SAIL) system.) At the end of the cruise the scientists naturally want to have a copy of their data.
In the past transferring cruise data was handled by writing out a 9 track tape. The scientist read the tape on their institution's main frame and that was the end of the problem. Actually sometimes it was the beginning of the problem, but this discussion will get long winded enough without getting to deep into obsolete technology.

Now the mainframe is disappearing and scientists are handling the data on a variety of workstations. Because 9 track tape drives are not very common on workstations the 9 track tape is no longer a suitable data transfer media.
As a result we have adopted several methods of transferring data, none of which are very satisfactory;

- If the scientist has brought his/her own computer, it can be connected to the network and the data can be copied directly to their computer. This assumes that the scientist will have adequate space on their machine and also a method of backing it up.
- Some scientists request their data be copied to IBM PC compatible floppy disks. This is always fun. Scientist, "Can I have my data now?" Technician,"lets see this cruise has been 20 days long that's 10 Mbyte per day times 20 days. Do you have 600 floppy disks with you?"

- What usually happens is that the scientist takes a subset of the data on floppies then calls back every few weeks asking for additional data-- a time waster on everybody's part.

Now that I have explained the situation I regularly find myself wrestling with, I will provide what I believe is "a" solution.

To have everyone using similar terminology I need to define some terms:
- ACSII Data Format (American Computer Standard for Information Interchange) - Files that contain characters that you can enter from a keyboard. Note that because Postscript files are made up of ASCII characters that tell a Postscript printer what to do, Postscript files are ASCII files.

- Binary Data Format - Non-ASCII, non-printable, non-interchangeable. To over simplify, the main problem here, byte ordering, most significant bit first (little endian) or most significant bit last (big endian).

- Native data format - Generally this is raw binary data. An example would be CTD data collected by General Oceanics Seasoft. The program only runs on a PC, the data collected by this software would only be post processed with a PC. The key idea is that this data is produced by a proprietary software package and must be further processed by the same proprietary software package before it is usable.

- Physically Compatible - In this case it is the physical dimensions of the storage media. A 3 1/2" floppy is physically compatible with PCs, Macs, etc.. In this discussion, it also means there is a reasonable way to connect a device using the specified physical size to all types of computers.

- Logically Compatible - In this case referring to sectors, tracks, directory structures, file naming conventions etc.. This is why you can't take a floppy written on a Mac and use it in a PC.

- Machine Independent - Totally independent of the manufacturer of the CPU and operating system you are using. This means physically compatible, logically compatible and binary compatible. This is what we want!!!!
- Small Data Sets - One to ten Mbytes of data. An example would be XBT drops. Not of much concern in this discussion because they are handled fairly easily.

- Intermediate Data Sets - Ten Mbytes to several GBytes. Raw ADCP data is typical of this magnitude.

- Large Data Sets - Several Gbytes to infinity. SEAbeam and multi-channel seismic systems are prime examples. Because almost all of this data is in a native data format and is processed on machines identical to the machine that it is collected on, this type of data is not included in this discussion.

- Self Describing File - Files which contain within itself a description of the data. For example the next 4 bytes is sea surface temperature in degrees Celsius based on the 1990 internatic standard to two decimal places and there are 2456 temperature readings taken at etc.. These files are generally binary files. Software subroutine calls are written in a common language such as "C" for each type of computer that needs to access the file.

- ISO 9660 - An international data standard that defines a file system for CD-ROMs. Almost all computer systems support CD-ROMS that conform to ISO 9660.

Primarily then, what I think is needed is a method of transferring

This is what I have summarized after reviewing the various types of data storage media presently available that may be used to transfer data to the scientist;

MEDIA	SIZE	PHYSICALLY	LOGICALLY	BINARY	Comment
Floppy	1-2M	yes	no	no	Too small
Floptical	20M	maybe	no	no	
9 Track	60M	maybe	maybe	no	
DAT	2-8G	yes	no	no	
8 mm	2-8G	yes	no	no	
MO	.3-.6G	yes	no	no	capacity/side
WORM	.5-1G	yes	no	no	
CD-ROM	.6G	yes	yes	no	ISO 9660

So, essentially the only media I know of that is both physically and logically machine independent is CD-ROM ISO 9660.

Hey, but aren't CD-ROMs those things that you get with all those games on them that they produce at some factory?

Well that's true but there is also a standard called CD-R which means CD recordable.

How do you record a CD?
Files that are to be placed on the CD-ROM are processed by a program that checks for proper file and directory names in accordanc with ISO 9660 standards. The files are then copied to a container file that is the same size as the CD-ROM to be recorded. The container file can then be used as a pseudo CD-ROM which can be be reviewed to determine that everything is as desired on the final CD-ROM. The container file is then copied to the CD-ROM. These steps are necessary because once the process of writing the CD-ROM is started it must be completed in one continuous sustained data transfer. The process of writing 600 MB to the CD-ROM takes 30 minutes on a double speed drive.

What do you need to record a CD?

A CD recording system consists of a computer (typically a PC), a CD-ROM recorder, a 700 MB hard disk (for creating a container file of the data to be placed on the CD), and software that is designed to create ISO 9660 files. Assuming you have a computer the rest of the goodies runs between $5000 and $8500. Blank CDs run around $23 each.

and he/she says "I don't have a CD-ROM reader." Tell the scientist

to go to their local discount computer store where they can purchase a CD-ROM reader for between $200 and $500. The $500 will buy a low end CD-ROM drive and a lot of CD-ROMs or a high end drive and no CD-ROMs., $200 will buy a low end drive and no CDs. It's still cheaper than 500 floppy disks and is a one time cost. Drives typically run higher for high end workstations (SUN, DEC, etc.) but many of these systems have CD-ROM drives anyway.

So, I have now described a distribution media that can be inserted into and read by just about any computer system out there. (Well, you probably can't read it on your Tandy 64 computer you bought in 1976.)

Wait a minute, where are we on machine independence? We are close but not there yet. The ISO 9660 standard says we can read the directory and ASCII files (including Postscript) that are on the disk, but we may not be able to read binary files because of endian problems.

There is one thing in our favor. Native data by definition will be processed on the same type of computer that wrote it. For example if the CTD data was collected using General Oceanics SEAsoft software on a PC, then any PC with a CD-ROM drive will be able to read and process the data using SEAsoft. So all we have left to deal with are binary files that needs to be handled on multiple platforms. Fortunately, standards have been developed that can solve this problem. Examples of two such standards are netCDF (network Common Data Format) and HDF (Hierarchical Data Format). Both of these standards use self describing files that are accessed though subroutine calls that have been developed for all of the major computer operating systems in use (UNIX, DOS, VMS etc.). Software written to access files written in these standards will be able to provide full machine independence. Several major software vendors have written software to access netCDF files.

Conclusion

CD-ROMs written in ISO 9660 format appear to go along way in solving the problem of transferring the data collected aboard RV Columbus Iselin to the scientists. (All of the data presently collected aboard RV Columbus Iselin is either in ASCII or native data format.) CD-ROM recording is now available at a reasonable cost. It provides physical and logical compatibility with most computers presently in use. It is projected not to deteriorate for twenty years.

I would like some input on this. If we all face a similar problem then possibly we can jointly develop a solution that we can all

I have only briefly mentioned standardized file formats such as netCDF and HDF. Standardized data formats provide a level of interchange that is extremely important in allowing scientists to readily use data from a variety of sources (in our case different ships in the UNOLS fleet). It is my hope that some other more qualified individual can start a discussion on this important subject.

P.S.
WARNING!!!! WARNING!!!

I don't own a CD-ROM recorder yet so I don't have first hand experience. I have spent a considerable amount of time talking to various venders etc.. I do have a SONY system coming for evaluation and will post additional information as I understand it. I drafted this discussion with the intent of providing actual experience with a CD-ROM recorder. I decided to post this now so that those who thought it worthwhile might include something in their instrument proposal to NSF.
The primary concern is whether these systems can be used at sea!

Sources for more information for internetters;

Usenet News groups:
comp.publish.cdrom.software

comp.publish.cdrom.hardware
sci.data.formats (for netCDF, HDF, other misc data formats)

Anonymous FTP

unidata.ucar.edu (for netCDF info)
ftp.ncsa.uiuc.edu (for info on HDF)

Where to get info or purchase CDR systems

Pinnacle Micro
19 Technoloy

Irvine, CA 92718
PH (800)553-7070

Approx. $4000
Advantages - lowest cost

Disadvantage - has small 64 K internal buffer, does not have double speed capability, not sutible for use on ship due to horizontal placement of +/- 5 degrees installation spec.

SONY CDW-900E

SONY Electronics Publishing Company
Publishers Data Service Corporation

One Lower Ragsdale Drive
Montery, CA 93940

TEL (408)372-2812
Approx. $6500

Advantages has 4 MB internal data buffer, ensures constant data transfer when recording, also has double speed capability.
Disadvantages - ?

Phillips CDD-521

Dataware Technologies
5775 Flatiron Parkway, Suite 220

Bolder, CO 80301
PH (303)449-4157

Approx. $6500
Has 256 kB buffer.

They also sell the SONY system for $7995
All prices include CD-ROM recorder, SCSI host adapter, ISO 9660 data formatting/recording software.

Start the Presses

Affordable do-it-yourself CD-ROM publishing will revolutionize how you distribute and use information

JON UDELL

When BYTE first featured CD-ROM on its cover in 1986, we said it would reinvent publishing. Today, CD-ROM is finally poised to deliver. CD-ROM drives are now faster and cheaper, and with the installed base (now estimated at 5 million) doubling every year, text-based and multimedia titles are proliferating. Kodak's Photo CD system establishes CD-ROM as a key ingredient in digital photography, and software makers are flocking to CD-ROM as a convenient alternative to distributing code and manuals. Businesses increasingly find CD-ROM a compelling medium for internal communication. And for governments obliged to distribute data to citizens, the medium of choice is becoming CD-ROM.

Into this bubbling mixture now drops a catalyst for even more rapid acceptance of CD-ROM: sub-$10,000 CD-R (CD recordable) drives that bring CD-ROM-making to the desktop. These new low-cost printing presses will shift the electronic publishing revolution into high gear.

CD-R is ideal for three kinds of applications: prototyping titles destined for conventional pressing, final production of discs for limited distribution or even single use, and archiving. Low-volume production is CD-R's forte.

At the National Library of Medicine (Bethesda, MD), a research library within the National Institutes of Health, searching for medical periodicals until recently meant wading through fat notebooks full of printouts. Now that data finds its way onto a custom CD-ROM. A single copy of the disc, produced monthly using the Sony CD-R drive and placed in a networked CD-ROM reader, gives researchers electronic access to medical citations.

Macess (Birmingham, AL), a developer of turnkey document-control systems for the managed health care industry, uses CD-R to store scanned images of claims forms. The discs reside in huge banks of networked CD-ROM readers, and Macess's software marries CD-ROM-based document images to fielded data in Btrieve databases kept on NetWare servers. One of Macess's clients scans more than 10,000 forms a day. When the accumulated image data reaches 600 MB—which happens once or twice a day—it's written to a CD-R disc using a Philips drive. The total capacity of the system will soon double to 168 networked readers and nearly 100 GB of document images.

The SAS Institute (Cary, NC), developer of the statistical software package that is called SPSS, uses the Makedisc software from Young Minds and the Philips

Affordable CD-R Drives

JON UDELL AND HOWARD EGLOWSTEIN

Once you read your first handmade CD-R disc in a standard CD-ROM player, you'll be hooked on the power of this exciting technology. Desktop CD-ROM recording is very new, though, and like any pioneer you should expect some of the hardships we encountered when we tested four of the current crop of CD-R solutions (see photo A). Their very different approaches to both hardware and software attest to the embryonic state of CD-R.

CD-Studio, YMI

We wondered how this deluxe solution for Unix-based CD-ROM makers could guarantee that a Unix host will sustain the 300 Kbps required by the double-speed Philips CDD 521 included with the package. The answer is that it doesn't have to. Young Minds, Inc., inserts a proprietary controller (shown in photo A) between the host and the recorder. The package sells for $18,250.

The CD-Studio controller is essentially a headless PC (i.e., no monitor or keyboard) with two high-performance SCSI adapters and a 1-GB disk. One adapter handles communications with the Unix host; the other talks to the Philips recorder. The controller powers up running custom YMI code that makes it appear to Unix as an 8-millimeter Exabyte tape device. To test that it's working, you can use the Unix `tar` command to copy files to the "tape" and read them back. What you can't do, though, is verify that the CD recorder is properly hooked up. We wish YMI had added LEDs to the controller's front panel to monitor the status of the recorder.

Our Unix test-bed was a Sun Sparcstation 2 running SunOS 4.1.2. YMI's `Makedisc` walked the tree of sources we fed it and built a 620-MB CD-ROM image on the controller's disk. To transfer that image to the recorder, we issued a standard Unix `mt` (mag tape) command, which tells the YMI controller's pseudo-tape device driver to start the recorder. After that, there's no feedback from the Unix console or the controller. You simply watch for the CDD 521's write LED to turn off.

Everything worked smoothly, and our first disc was a success, although even BYTE's Unix experts found YMI's documentation cryptic. Because we used `Makedisc`'s -R option to add Rock Ridge extensions, the resulting CD-ROM had the look of a Posix file system (e.g., long names, symbolic links, and permissions) when mounted in the Sparcstation's CD-ROM drive. (DOS and Macintosh systems saw the same disc as a standard ISO 9660 CD-ROM.)

`Makedisc` also stores descriptions of the Unix-to-ISO 9660 mapping in a text file at each node of the ISO 9660 tree. That way, Unixes other than the two that now support Rock Ridge (SunOS and NextStep) can recover the original names. YMI's `cd_link` enables such systems to map a hard disk-based tree of real Unix filenames to the corresponding tree of short, uppercase-only ISO 9660 filenames on the CD-ROM. YMI also provides an NFS (Network File System)-compatible file system called PFS (Portable File System) that handles CD-ROM sharing over Unix networks more efficiently than NFS.

Personal RomMaker, JVC

Mastering your disc on JVC Information Products' Personal RomMaker is about as easy as it gets. Personal RomMaker ($12,799) requires a Mac IIci or better, 4 MB of RAM, System 7.x, and HyperCard 2.1 or higher. In addition, you'll need enough disk space to store your original data files, but there's no need for fast storage here: RomMaker can run happily off shared disks or network servers.

To burn a ROM, you need either fast access to the data files or a fast hard drive with a premastered image. JVC uses the latter and incorporates a high-speed Maxtor 8760 SCSI drive into the Personal RomMaker system (this is the same drive we chose for our DOS testing). You create your disc by deciding what files you want stored and how you want them arranged. RomMaker gives you a simple script editor for creating an ASCII list of these files and folders. We'd prefer it if RomMaker used standard Apple file dialogues for selecting files, but the ASCII file syntax is simple enough to handle.

The premastering stage runs through your file list and then copies each file from the Mac's file system to RomMaker's hard drive in CD-ROM format. You can choose from several disc formats: Apple HFS, ISO 9660, High

Photo A: *(Clockwise from bottom left): The Sony CDW-900E, Philips CDD 521 (in Kodak guise), JVC Personal RomMaker, and YMI CD-Studio controller.*

Sierra, or a hybrid HFS and ISO 9660 format. Our test disc consists of 620 MB of images. To find enough free disk space, we used System 7's file sharing to convert one of our Quadras into a file server. Over a thin Ethernet link, it took about 1¼ hours to suck the 620 MB through the network and create the CD-ROM image.

After the premastering, you can mount the hard drive as a CD-ROM image, manipulate the folder and window structure, and test your application. To create the CD-ROM, you simply choose the appropriate menu selection. After verifying that you've placed a blank in the drive, RomMaker trundles off and copies the hard drive contents to the CD-ROM recorder. If you have the optional DAT (digital audiotape) drive, you can opt to copy the image to DAT for a commercial premastering house.

CD Record, Dataware

Dataware resells the Philips CDD 521 for the DOS market with its CD Record software ($8995 for the package). Unlike Philips' CD-Write, which performs ISO 9660 formatting on the fly as it transfers files from a hard disk to the write-once disc, CD Record requires two steps: First build a virtual image, and then write that image to the CD.

First-generation premastering systems, like Dataware's CD Prepare, convert a DOS hierarchy that may contain thousands of files into one huge file that's a bit-for-bit image of an ISO 9660 file system. The image file serves two purposes. With the software emulator such systems typically provide, you can fool MSCDEX into seeing the image as a real CD-ROM. That means you can test your retrieval application before mastering any discs. Once satisfied with the image, you dump it to tape for shipment to a mastering house.

CD-R changes the rules of the game. Now you can build your own test disc, which is also a convenient delivery vehicle. In principle, there's no need for an image file, or for an extra 660-MB hard disk to store it on. But given the high cost of CD-R media, it's still useful to have an image file available for preliminary testing. CD Record's virtual image meets that need—without doubling your storage requirement—by adding a thin ISO 9660 mapping layer to an existing DOS file system.

CD Record has some drawbacks, though. Building that image took us over an hour for our 8000-file data set, and that still left an additional half hour to cut the disc. By contrast, CD-Write, although lacking the useful virtual image capability, does the entire job—formatting and disc cutting—in half an hour. Also, CD Record's virtual image lives in RAM and does not survive a reboot. If you need to make a CONFIG.SYS change after running `vmap` (the ISO 9660 mapper) but before starting `cdrecord` (the recorder), you'll have to rerun `vmap`—an inconvenience that Dataware admits should be fixed.

Hardware trouble botched our first two discs. While DOS can disgorge files fast enough to sustain 300 Kbps to the double-speed CDD 521 for the duration of a write session, conditions must be near-optimal. After rearranging some questionable SCSI cabling, we did successfully cut discs. But the experience reminded us that a more expensive solution involving a dedicated hard disk (e.g., YMI, JVC) is also inherently more reliable.

Multimedia Formatter, Sony

Sony's rack-mountable CDW-900E (not sold at retail) offers several unique features. In addition to the SCSI-2 connectors, its back panel sports an extra pair of proprietary connectors. These enable you to link a master CDW-900E to as many as 15 slaves for simultaneously writing multiple CDs. (Additional connectors, not currently used in any PC-hosted application, support recording audio CD.) The drive's capacious 3-MB buffer helps ensure the sustained flow of data on which CD recording critically depends. Moreover, since it can run at single or double speed, the host need not support the 300 Kbps required by the Philips drive.

Like CD Record, Sony's Multimedia ISO Formatter layers a virtual image on top of an existing DOS file system. However, it provides no emulator, nor can the drive work as an ordinary CD-ROM reader (as the Philips drive does). When you cut a disc from a virtual image, Sony runs the drive at single speed to ensure best results. For faster writing, you have the option of building a real image on the hard disk and cutting it to the disc at double speed.

Like the drive itself, Sony's formatting software has some unique features. In addition to straight ISO 9660 CD-ROMs, it can apply the secret sauce needed to make titles for Sony's Data Discman and MMCD Player. The Data Discman plays 8-centimeter discs that are in XA format but have nonstandard volume information. Another secret recipe makes the MMCD Player's regular-size XA discs. Sony's formatting software, which presents a DOS-based GUI, supports both these formats.

For Unix users, YMI's CD-Studio, with its Rock Ridge support, is a solid—albeit pricey—solution. Mac users who can afford JVC's Personal RomMaker will appreciate its outstanding ease of use and HFS support. For DOS users, both the Sony and Philips drives have their advantages. The Sony CDW-900E's large buffer and single-speed mode could improve reliability in some configurations, and its support for Sony's Data Discman and MMCD Player is unique. The Philips drive, widely supported by resellers, is a solid performer and—at least for now—the most inexpensive CD-R solution.

Jon Udell is a BYTE senior technical editor at large. Howard Eglowstein is a BYTE Lab testing editor. You can contact them on BIX as "judell" and "heglowstein," respectively.

COMPANY INFO

Dataware Technologies, Inc.
(617) 621-0820
fax: (617) 621-0307
Circle 1075 on Inquiry Card.

JVC Information Products Co.
(714) 965-2610
fax: (714) 968-9071
Circle 1076 on Inquiry Card.

Sony Corp. of America
(800) 352-7669
(201) 930-6432
Circle 1077 on Inquiry Card.

Young Minds, Inc.
(909) 335-1350
fax: (909) 798-0488
Circle 1078 on Inquiry Card.

Buying a CD-ROM Drive

TOM HALFHILL

The proliferation of CD standards, including CD-ROM XA and multisession Photo CD, makes it harder than ever to decide which drive to buy. Further complicating the decision is the availability of dual-speed drives that, for a $200 premium, double throughput. One alternative is simply to wait a few months: Drive makers are constantly updating their products, and they say that adding XA or Photo CD support won't be a big deal. But the definitions of these features can be slippery.

Consider NEC's "XA-ready" drives, which can read mode 2 format. To be fully XA compatible, they will need additional circuitry to decode interleaved channels of audio and perform ADPCM (Adaptive Differential Pulse Code Modulation) decompression. To add these capabilities, you're likely to need a new interface card. The same is true of XA-ready drives made by Sony and others. Currently, the only Sony devices with built-in ADPCM chips are the Data Discman and the MMCD Player. Sony says, however, that it will add the relatively inexpensive ADPCM chips to its CD-ROM drives if XA compatibility becomes more important.

So far, the most compelling reason to have XA is Kodak's Photo CD. To read a Photo CD, a CD-ROM drive must be XA ready, although it does not have to be fully XA compatible. However, full XA compatibility with ADPCM will be required for future Photo CD applications that will use interleaved audio. Lacking ADPCM, an XA-ready CD-ROM drive could play sound bites attached to individual images, but it couldn't play a continuous audio track while reading one image after another. For an impressive business presentation or narrated slide show, you'd need full XA.

Another desirable feature in a CD-ROM drive is the ability to read multisession Photo CDs. Several drives can read the first batch of images written to a Photo CD disc. But when you add another batch of pictures, it becomes a multisession CD, and only a handful of drives can track the modified directory structure and locate the additional images.

"I don't think multisession is a big deal," says Pat Fobes, a CD-ROM hardware manager at NEC Technologies (Wood Dale, IL), "but Kodak is putting a lot of marketing behind the idea of multisession, and the public perception will be that multisession is important." As a result, virtually all major manufacturers will be adding multisession capability to their new drives. Usually it requires a patch to the drive's firmware, plus some modifications to the servo tracking system so the device knows what to do when the laser pickup stumbles on an unrecorded region of the disc. Such modifications add little or nothing to the retail price of a CD-ROM drive.

If you opt for a dual-speed drive, though, you will pay a premium. NEC introduced this feature in 1992, and it's catching on fast. To see why, compare the specifications: A typical single-speed drive might have an average seek time of 450 milliseconds and a data transfer rate of 150 Kbps. A dual-speed drive's numbers might be 280 ms and 300 Kbps. Such performance is still anemic by hard drive standards, but it makes a big difference when you're reading high-resolution Photo CD images, which can run to 6 MB in size. The next frontier is likely to be 600-Kbps "quadspeed" drives like Pioneer's new DRM-604X.

Here's the bottom line. If you need a CD-ROM drive to access static information—encyclopedias, technical manuals, reference books, and so on—you can get by with an inexpensive single-speed drive without XA or multisession support. For multimedia CD-ROM applications involving sound and animation, consider a dual-speed drive for best throughput. If you anticipate using Photo CD at all, you'll need at least an XA-ready single-session drive. For serious Photo CD work, settle for nothing less than a dual-speed drive with full XA and multisession capability.

Tom Halfhill is BYTE's senior news editor in San Francisco. You can reach him on BIX as "thalfhill."

---

drive to build custom CD-ROMs for its Unix clientele. For each disc, SAS selects an appropriate subset of programs from its large family of products and uniquely serializes the programs to copy-protect them.

The U.S. Geological Survey (Reston, VA) has used CD-R since 1989. To equip its field offices, a $60,000 Meridian solution that uses a Yamaha drive "was just too rich for the taxpayers' pocketbook," says Dave Traudt, manager of the USGS CD-ROM support center. Traudt placed low-cost CD-R drives in seven field offices and is buying four more. "Our agency has a very tight budget; that should tell you a lot about how we see recordable technology," he notes.

CD-R has been available to businesses flush with cash for several years. Now, affordable CD-R from Philips (the $7995 CDD 521), Sony (the CD-900W, not sold at retail), JVC (the $12,799 Personal RomMaker and $9995 Personal Archiver), and Pinnacle Micro (the $4995 RCD 202) have put the technology within reach of a vast new segment of users (see the text box "Affordable CD-R Drives" on page 118).

One Drive Reads All

Without CD-R, CD-ROM mastering and replication require that the data, usually on 9-track tape or DAT (digital audiotape), be sent to an outside duplicator that uses the same multimillion-dollar equipment that stamps out audio CDs. The new recordable drives produce an equivalent product—CD-ROM discs containing files in the standard ISO 9660 format that ordinary CD-ROM players can read (see the text box "Buying a CD-ROM Drive" above). But CD-R (or CD Write-Once) discs aren't replicated from a master. Each is a custom edition that is made by copying data from a hard disk to a specially made blank disc.

The CD-R drive's high-powered laser burns pits into a pregrooved 120-millimeter disc that is a sandwich of polycarbonate substrate, organic dye, and a gold reflective layer (see figures 1 and 2). The half-hour or hour required for this transfer and the $25 to $40 cost of the blank dictate that CD-R will complement rather than replace mastering and replication.

Popular Publishing

All organizations collect and disseminate information. Several innovative ones now publish their parts catalogs, training manuals, technical documentation, and other forms of mission-critical data on CD-ROM. A CD-ROM disc is not only infinitely smaller, lighter, and cheaper to transmit than the equivalent tower of 200 Webster's dictionaries, but is also far more useful, because it is electronically searchable.

By lowering the barrier to entry, CD-R will dramatically expand the number of CD-ROM publishers and applications. Companies that produce discs using CD-R have to worry about the same duplication economics as before, to a point. Depending on the cost of the CD-R media, the break-even point for making copies using CD-R is between 20 and 100 discs. Above that, conventional replication will be cheaper (see figure 3 on page 131). In some cases, factors that aren't purely economic (e.g., convenience or security) could push the ball into CD-R's court.

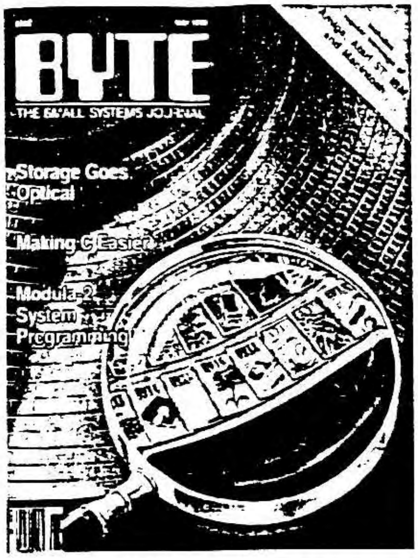


Photo 1: *BYTE touted the promise of CD-ROM to revolutionize publishing on the cover of our May 1986 issue. Today, CD-R appears poised to deliver on that promise.*

While reviewing the CDD 521 Philips drive (see "Desktop CD-ROM Publishing," January BYTE), I built a prototype of a disc containing BYTE's text from the last six years—a possibility suggested by the BYTE cover story on CD-ROM in the May 1986 issue (see photo 1). (The project had long been contemplated as a tool for editorial research.) The recordable drive made the project feasible, and being able to demonstrate a real CD-ROM disc galvanized support as nothing else could have.

I doubt that our situation is unique. CD-R's advent will spur countless fence-sitters to action. Says Larry Schiller, president and CEO of the Bureau of Electronic Publishing, a CD-ROM distributor (Parsippany, NJ), "Just about every business with sales of $10 million or more will buy a recording drive."

Internal disc production has security benefits, too. Military, government, or corporate users who must publish highly sensitive information on CD-ROM may not want to release that data, even temporarily, to an outside mastering facility. With a secure in-house CD-R system, the secrets can be safely reproduced.

continued

STANDARD VS. CD-R DRIVE LASER MECHANISM

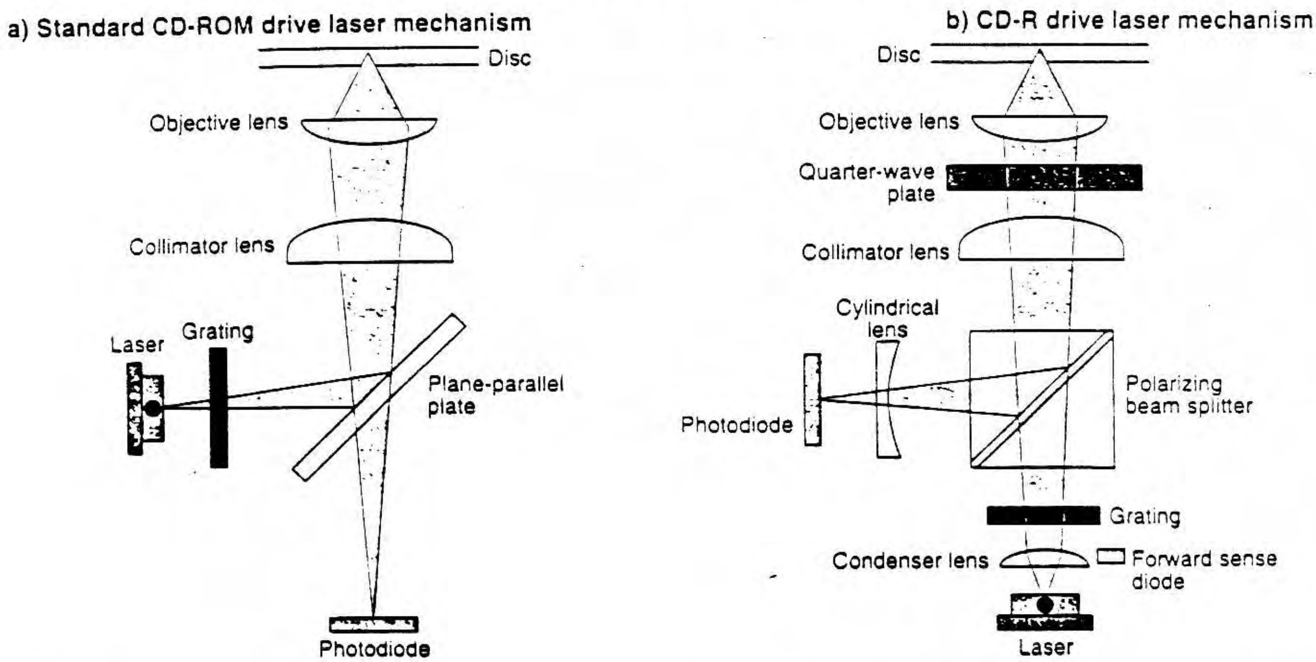


Figure 1: (a) *In a standard CD-ROM drive, the laser's orientation is perpendicular. The beam takes a 90-degree bend en route to the disc and then reflects back to the photodiode.* **(b)** *In a CD-R drive, the path from laser to disc runs straight. The condenser lens couples the light to improve the laser's output, and the forward sense diode measures its intensity. The quarter-wave plate rotates the beam's polarizing direction twice, so the reflected light takes a 90-degree bend on its way to the photodiode. (Figures courtesy of Philips Consumer Electronics Co.)*

The New Breed of CD Players

ED PERRATORE

Three unusual CD players debuted in 1992: Sony's MMCD Player, Tandy's VIS (Video Information System), and Philips' CDI 360 Portable. Each uses a different processor, operating system, and CD format (see the table). The common threads are CD-ROM and NTSC output. These three specialized CD appliances are all equipped to play on your TV.

The VIS, restricted to that mode, most clearly targets the home market. VIS applications use a Windows derivative called Modular Windows, which presents a simplified interface on a TV screen and takes user input from an infrared remote control. Developers will use familiar Windows tools to create VIS applications and can readily port MPC titles to the new platform. Owl International (Bellevue, WA) and AimTech (Nashua, NH) have announced authoring tools to support the VIS.

Initially, the VIS CD-ROM drive supports neither single-session nor multisession Photo CD. However, the VIS is eminently upgradable, with sockets for a video accelerator and a modem. "It's the most open-architecture computer out there," says Richard Doherty, editor in chief of the multimedia newsletter *Envisioneering* (Seaford, NY), "but Tandy will say, No! No! It's not a computer, it's an appliance!"

The Sony and Philips players, which can connect to a TV or run stand-alone, double as entertainment or educational appliances and as business tools. *Newsweek*'s recently announced *Newsweek Interactive*, a quarterly CD-ROM version of the magazine, will showcase the MMCD Player as a home entertainment device. Northern Telecom's new diagnostic application for PBX systems will put the MMCD Player to work. The company's 3700 field technicians will soon be able to connect the Sony unit's serial port to PBX switches, run diagnostic software, and hot-key to a 30,000-page store of searchable documentation.

Philips' CDI 360 can also play in both consumer and business markets. The company is marketing CD-I as a better platform than CD-ROM XA for integration of audio, video, text, and graphics. An alliance between Dataware Technologies and Philips subsidiary OptImage brings the data-crunching capability of trac ional CD-ROM retrieval software to the sight-and-sound world of CD-I.

Although the MMCD Player's use of the CD-ROM XA format theoretically enables it to share titles with ordinary PCs that have XA drives, there are obstacles. Most XA drives sold today don't come with the hardware support (ADPCM [Adaptive Differential Pulse Code Modulation] decompression and deinterleaving) that Sony's embedded XA chip set provides. Also, the Sony unit's small display and lack of a hard drive make it incompatible with current XA-based PC applications.

For Philips, the cost of CD-I hardware and development tools is an impediment, as is the royalty the company charges for use of the CD-I name. "When you say 'CD-I,' you are mentally writing a check to Eindhoven [the Netherlands base of Philips Electronics NV]," says Doherty, "and that rubs a lot of people the wrong way."

CD-ROM DRIVES FOR YOUR TV

The VIS, with no display of its own, allows output options to near-VGA 640- by 400-pixel and lower resolutions in addition to NTSC. Part of the CDI 360's price is for a 6-inch active-matrix color display unmatched by the MMCD. (N/A = not applicable.)

	CDI 360	MMCD	VIS
Processor	68000	8086	286
Operating system	OS/9	DOS	Modular Windows
Disc format	CD-I	CD-ROM XA	CD-ROM
RAM	8 KB nonvolume	640 KB	1 MB
Attached display	Yes	Yes	No
Serial port	Yes	Yes	No
Display resolution (pixels)	756 × 556	300 × 200	N/A
Compressed audio in hardware	Yes	Yes	No
Full-motion video in hardware	Available first quarter 1993	No	No
NTSC output	Yes	Yes	Yes
PAL output	Yes	No	No
Portable	Yes	Yes	No
Price	$2000	$1000	$700

Ed Perratore is a BYTE news editor based in New York. You can reach him on BIX as "eperratore."

One-of-a-kind discs could also be useful business development and sales tools, says Rich Bowers, executive director of the OPA (Optical Publishing Association) (Columbus, OH). Each year, 100,000 companies bid to sell their wares to the U.S. government, and a response to a federal request for proposals can run to hundreds of pages. Similarly, pharmaceutical companies require heaps of documentation to obtain FDA approval for new drugs. Submitting the required 20 copies of your proposal on CD-ROM, says Bowers, is easier for you to send and for the recipient to evaluate. CD-ROM also gives you the option of making a more

CD-R MEDIA COMPOSITION

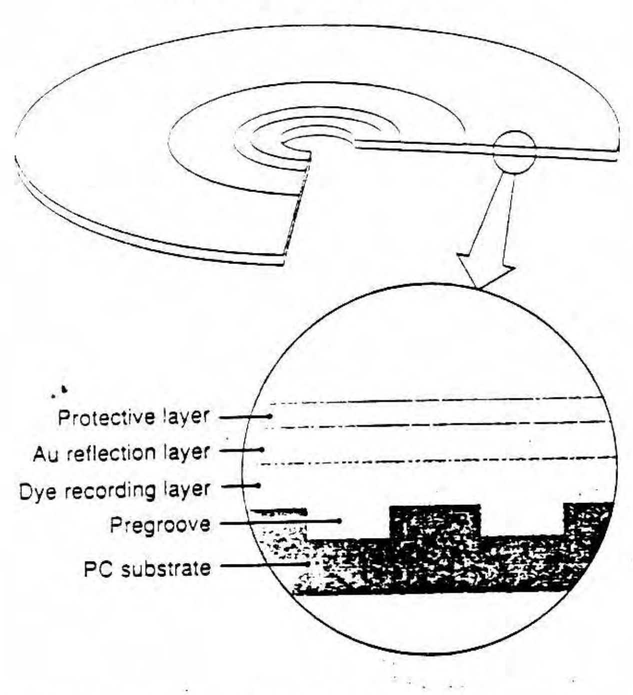


Figure 2: *The CD-R media consist of four layers: the top protective layer, the gold (Au) reflection layer, the dye recording layer, and the PC substrate. The CD-R laser heats the reflection, recording, and substrate layers. As a result, the recording layer fuses and is impressed by the expanding substrate layer. (Figure courtesy of Philips Consumer Electronics Co.)*

impressive sales pitch by jazzing up the content with sound and video.

For governments, a compelling use of CD-R will be to satisfy requests for information. Conventional CD-ROM publishing has already improved matters tremendously. Topographical and geological survey data used to cost $100 per 9-track tape. Now you can get it on CD-ROM for $32, says George Knapp, a USGS abstracts editor and chairman of SIGSOFT, a group that evaluates CD-ROM authoring and retrieval tools. "For public access to data," he says, "CD-ROM is the greatest thing that's ever happened."

The Office of Business Analysis in the U.S. Department of Commerce (Washington, DC) publishes the National Trade Data Bank on a CD-ROM that amalgamates import/export data, marketing studies, and foreign-aid program information from 15 federal agencies. The disc helps U.S. exporters understand international markets. A second title that delivers U.S. economic, social, and environmental data will soon be available.

CD-ROM discs produced under federal auspices are almost always replicated, even if they're ultimately sold only in small numbers, because copies must be distributed to the federal depository libraries—a group of some 1250 university and public libraries that agree to store federal data issued through the U.S. Government Printing Office. Nevertheless, there is a need to publish individual discs on demand.

Daniel Costanzo, a physical scientist with the U.S. Army Topographic Engineering Center (Ft. Belvoir, VA), expects that customized slices of topographical data that today ship on 9-track tape will eventually be delivered on CD-ROM discs. There's also an argument in favor of publishing some kinds of widely distributed data on demand. "If you print 10,000 CDs of GIS [geographic information system] data, it's like printing 10,000 maps," he says; "the currency deteriorates over time." In cases where demand-published CDs are feasible, users will always get fresh data.

The nature of that data will determine how well such a model works. Satellite and sonar images, which are just collections of files, are easy to dump to a disc. In the case of text or fielded data, however, the need to index the material will make it harder to carve out and deliver unique subsets.

Publisher Profiles

For CD-ROM publishers, testing discs prior to mass production was a complex ritual before CD-R. An emulator like Meridian's CD Publisher could make a hard disk–based ISO 9660 "image" written by an authoring tool look like a CD-ROM to the retrieval software. Such emulators can artificially slow access to the hard disk to help you gauge what the actual CD-ROM performance will be.

But one person sitting at the emulator's console can't fully explore the vastness of a disc and ensure that hundreds of megabytes of data have been properly assembled. For that, you need to make discs that can be distributed to a team of testers equipped with ordinary, inexpensive players. Suppose you need 10 such discs. Producing them by conventional means might entail a minimum order of 100, cost $1500, and take three to five days. The CD-R solution, by contrast, yields just the 10 discs you need, for about $300, in one day. "After you've gone around that track a few times," says Bill Harlow, a marketing manager at Philips Consumer Electronics (Knoxville, TN), "our drive has already paid for itself." On a tight deadline, the rapid turnaround of CD-R—for which mastering houses may double their fees—can be a critical factor.

The number of CD-ROM publishers is growing rapidly. Market researcher InfoTech (Woodstock, VT), which has tracked the CD-ROM industry since its inception, reports that last year the number of commercial and in-house titles grew from 3500 to 5000, and the installed base of drives more than doubled, to over 5 million (see figure 4 on page 132). At the same time, the CD-ROM retail market experienced dramatic change, according to the Bureau of Electronic Publishing. In 1991, fewer than 20 percent of the 17,000 computer-related retail outlets in the U.S. carried CD-ROM products. Last year, 80 percent did.

The Library Corp. (Inwood, WV) helps public and university libraries merge their own collections data with general sources such as indexes of periodicals, thereby creating custom titles unique to each library. Conventional replication is the technique of choice for bigger customers, such as the St. Louis Public Library, which has 141 networked readers. But CD-R is The Library Corp.'s ace in the hole for clients like Texas Christian University, which needs only a single disc.

Donnelly Marketing Information Services (Stamford, CT) makes CD-ROMs for market researchers. DMIS's director of software development, Mike Herman, says his company is investigating ways to create "onesie-twosie" discs for clients who want to merge their own data with the DMIS data. For Herman, the human cost of producing custom discs is by far the biggest issue.

At Hub Data (Cambridge, MA), a publisher of financial data on CD-ROM, CEO Bob Huebscher views CD-R as a way to deliver incremental updates to his monthly product. Huebscher sees the cost of CD-R blanks as a bigger obstacle. "I wouldn't think twice about a $2 disc," he says, "and I'd consider $10, but the current

Once and Future Standards

A flurry of emerging standards swirls around CD-ROM and CD-R. For starters, there's a new "color book" on the horizon. The Crayola box of CD standards (see figure A) has for some years included Red Book (CD audio), Yellow Book (CD-ROM), and Green Book (CD-I). (CD-ROM XA doesn't rate its own color; it works within Yellow Book's free-form mode 2 sectors.) Now there's Orange Book, which governs MO (magneto-optical) and CD-R drives. Part 1 of Orange Book defines a CD-ROM/MO hybrid. Part 2 defines single-session and multisession CD-R, and it is the subject of much current confusion.

Because the specification isn't nailed down, true Orange Book devices don't exist yet. Among CD-R drives, the Philips CDD 521 follows Orange Book most closely, although Philips admits a drive bought today might need a firmware upgrade next year. Sony has chosen to wait for a final specification before building multisession support into its CDW-900E. Alan Sund, Sony's (San Jose, CA) marketing manager for CD-ROM drives, says, "We're not going to have a compromised product claiming to follow a standard that doesn't yet exist."

Meanwhile, vendors of CD-ROM players—including Pioneer, Toshiba, Philips, and, yes, Sony—are scrambling to make their drives work with multisession Photo CD discs. Kodak, too, is betting that its own Photo CD player won't become obsolete by some last-minute change in the specification.

Why the confusion over multisession CD? Early developers of CD-ROM never anticipated CD-R. Users would read CD-ROMs, not write them, and a single session that could hold oceans of data hardly seemed limiting. (In retrospect, the handwriting was on the wall. From the beginning there were "mixed-mode" discs that combined file-system data and audio tracks.)

Those same assumptions led to the ISO 9660 file system, which is both a great strength of CD-ROM and an increasingly troublesome limitation. ISO 9660's strength lies in its unique status as an operating-system-independent file system. An ISO 9660–formatted disc works identically on a PC, a Mac, or a Sun workstation. The disc may include operating-system-specific versions of a retrieval application, but each of these will access a common set of files. However, ISO 9660 is a least-common-denominator file system: It sacrifices features of the Mac (e.g., icons and resource forks) and Unix (e.g., symbolic links and permissions), and it doesn't permit updates.

There are two ways to make richer CD-ROM file systems. One is to make a non–ISO 9660 CD that uses a native Macintosh or Unix file system. That works, but performance can suffer since these file systems require special tuning for slow media. Also, you lose the vaunted interoperability of CD-ROM.

A second approach is to wrap operating-system-specific extensions around an ISO 9660 core. The Unix version of this technique is the Rock Ridge protocol, a compatible superset of ISO 9660. Unix systems equipped with the Rock Ridge extensions see a CD-ROM in Rock Ridge format as a Posix file system, complete with long filenames, permissions, and symbolic links. Other systems see the same collection of files as an ordinary ISO 9660 namespace. JVC Information Products' (Huntington Beach, CA) Personal RomMaker applies the same concept to the Macintosh. Its hybrid HFS/ISO 9660 mode makes discs that look like Macintosh volumes to Mac users but work for DOS and Unix users as well.

On the drawing board is a specification for a new CD-ROM file system. Known informally as the Frankfurt specification (officially, standard 168 of the ECMA [European Computer Manufacturers' Association]), it promises equal enrichment for Unix, Mac, OS/2, and Windows NT. Working hand-in-hand with Orange Book, Frankfurt will also support the incremental update capability that ISO 9660's designers never thought would be needed. The Orange Book/Frankfurt combination promises to make tomorrow's CD-R systems as easy to update as today's WORMs, while elevating the cross-platform CD-ROM standard to meet the needs of modern operating systems.

The catch? Everything breaks. Drives will need firmware upgrades, and no current CD-ROM software infrastructure will carry over to Frankfurt discs. That's why Young Minds' (Redlands, CA) president Andrew Young says of the Rock Ridge extensions he invented: "We're solving a today problem. Frankfurt solves a tomorrow problem."

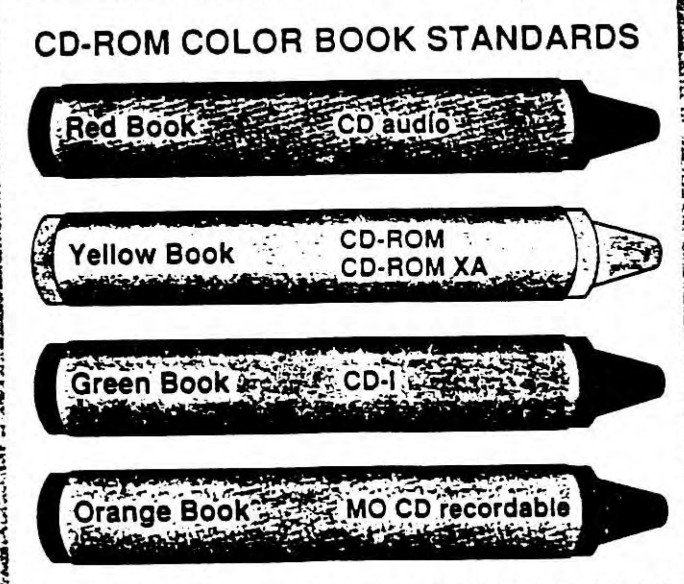


Figure A: *CD-ROM standard specifications are categorized by color. The new color is Orange, which covers MO and CD-R.*

$25 just isn't in the ballpark." The frequency of publication combined with high media cost would shrink Hub Data's profit margins.

Not much can be done about the labor cost of CD-R production, although some drives (e.g., Sony's) can be used in parallel to produce multiple copies at a time. Media costs are falling, however; the only questions are how far and how fast. Eastman Kodak (Rochester, NY) predicts $10 blanks in a few years. Is that likely? Much depends on the success of Kodak's Photo CD, a new imaging process that is the granddaddy of all CD-R applications. (Kodak writes Photo CD discs using the Philips CDD 521 drive.) Photo CD could be the engine that drives CD-R, just as audio CD powered CD-ROM. Moving just 1 percent of the film images processed each year into Photo CD format would create 5 million custom discs.

CD-ROM has also become a popular vehicle for software distribution. "We don't have final numbers yet," says InfoTech researcher Deborah Barlow, "but I won't be surprised if the number of discs used for software delivery exceeds those published for sale as commercial titles." Sun Microsystems, Hewlett-Packard, and Apple have long used CD-ROM to deliver software, but this year has seen many new entrants, including IBM (OS/2), Microsoft (Windows NT), and Borland International (Borland C++). Windows NT alone, still in prerelease, has already shipped on CD-ROM to 20,000 software developers.

The U.S. government is keeping pace with this trend as well. Allan Betts, CD-ROM production chief at the National Technical Information Service (Springfield, VA), says that the U.S. government produced 2000 titles last year, and he predicts rapid future growth. Betts, who advises dozens of federal agencies on CD-ROM production, says that one of the pressing plants his clients use now gets more than half its input in the form of CD-R discs rather than the traditional magnetic tapes.

Another segment builds multimedia titles for MPC, CD-ROM XA (Extended Architecture), and CD-I (Compact Disc Interactive) platforms (see the text box "The New Breed of CD Players" on page 124). The authoring tools these publishers use differ from the ones that create conventional titles, especially when the target is an XA or CD-I player that expects a mode 2 format incompatible with the mode 1 format of conventional CD-ROM (see the text box "Once and Future Standards" on page 130). But these distinctions exist purely at the logical level. To a CD-R drive, it's all just bits. Users of Mammoth Micro Productions' (Golden, CO) Studio/XA or OptImage's (West Des Moines, IA) MediaMogul test their XA and CD-I titles on the same CD-R drives that Dataware Technologies (Cambridge, MA) or SilverPlatter (Norwood, MA) customers use to develop standard CD-ROMs.

Manage Your Assets

As CD-ROM technology grows more practical, companies often find that they're unprepared to meet the challenge of what the OPA's Bowers calls "enterprise publishing." According to Bowers, CD-ROM has catalyzed a whole new view of the corporate information asset. "When you shift to this view," he says, "you suddenly find your whole structure [the way corporate data is stored] is upside down." Text kept in a random assortment of formats, tagged haphazardly or not at all, won't easily flow into a structured, possibly hyperlinked CD-ROM application.

Vendors of authoring and retrieval tools—including Fulcrum, Verity, and Personal Library Software—use fuzzy search and relevancy techniques, as opposed to standard Boolean searches, to minimize the need for structure. Fuzzy searches look for an approximate match; for example, a search for the word *war* might also turn up references to *conflict* or *battle*. Relevancy examines the context of the retrieved text to determine the most appropriate hits.

Such products also support fields, however. Structure married to sophisticated full-text search capability is the most powerful combination. Electronic Book Technologies' (Providence, RI) DynaText exploits that combination in a striking way: When you do a search, it distributes the hits across the table of contents, so you can see at a glance which are the relevant chapters or sections. Thus, instead of getting hits on a per-document basis, you see how they relate to structure internal to documents.

How do you put in the structure? A consensus is rapidly emerging: SGML (Standard Generalized Markup Language). SGML is an extensible system for describing the structure and style of richly formatted documents (see "SGML Frees Information," June 1992 BYTE). Like many government-mandated standards, it bores the average person to tears. However, it is clearly the right way to manage text for dual use—that is, for simultaneous print and electronic publication. That's why both Novell and Silicon Graphics recently decided to migrate their documentation from proprietary formats to SGML. Avalanche Development (Boulder, CO) and Exoterica (Ottawa, Ontario, Canada) offer SGML conversion tools. Dataware Technologies' CD Hypertext, which natively comprehends SGML, is an authoring tool that the company used to build a hyperlinked CD-ROM version of the U.S. tax code.

Another strategy for text is Adobe's Acrobat. Acrobat builds on two core technologies: Adobe's Multiple Master fonts, which can emulate the metrics of virtually any font, and a "reduced instruction set" PostScript that compresses files to a fraction of their normal size. The architecture includes a file standard, known as PDF (Portable Document Format), and software to convert

CD-R VS. CONVENTIONAL CD PRODUCTION

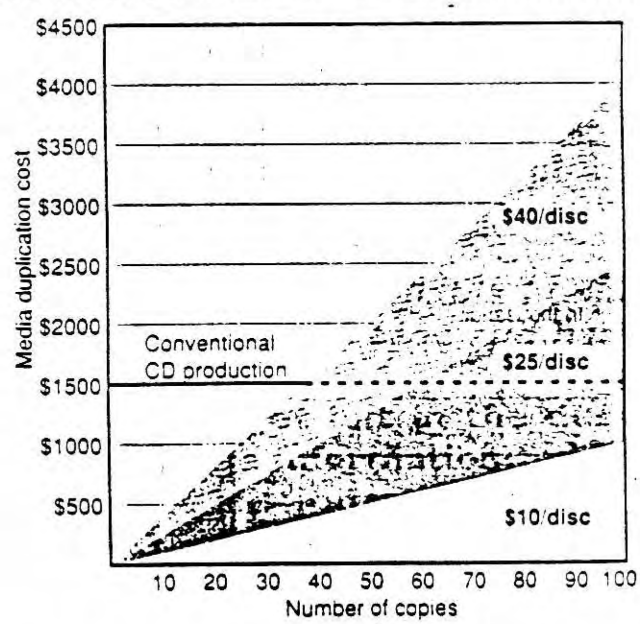


Figure 3: *CD-R production pays as long as the number of copies stays relatively low. Conventional CD-ROM duplicators charge about $1500 for 100 discs; that rate can double for one-day turnaround. At $40 per disc, CD-R is cost effective for press runs of under 35. This chart does not take into account the cost of the CD-R drive or the cost of labor.*

PostScript and non-PostScript documents into PDF format. Inexpensive software viewers that allow you to read and navigate PDF documents will be supplied first for Windows PCs and Macs and later for Unix, DOS, and OS/2 systems.

Using Acrobat, you can assemble documents created in a desktop publishing program such as PageMaker and run them through a batch Distiller that converts them to PDFs. Or you can use the PDF Writer to convert non-PostScript word processing files from Word or WordPerfect. Once these documents are written to a CD-ROM or other distribution medium, any system that hosts a viewer application can read them. In fact, viewers customized to read only the documents on a given CD-ROM can even be shipped with each disk. Note, though, that the first release of Acrobat won't support full text indexing.

The correct dual-use strategy for nontextual data is far less obvious. You can't just dump a Clipper or FoxPro application onto a CD-ROM disc and expect anything close to reasonable performance against large data sets. While the newer CD players can transfer data at a respectable clip, their slow seek times relative to those of hard disks remain a deadly impediment. Commercial tools like Dataware's CD Author use a variety of tricks to optimize for CD-ROM: loading B-trees fully (since they can't grow in a read-only environment), storing indexes as separate files, and exploiting disc geography for best locality of reference. With these techniques, a simple search can take just seconds instead of many minutes.

Complex queries present even trickier problems, however. A SQL query that wanders all over the disc will take a year and a day, and it has no place to write its temporary indexes. Many producers of CD-ROM database products take a radically asymmetrical approach: They anticipate common queries, generate result sets in advance (often represented as bit maps), and write the prepared results to the disc alongside the data. Such preparation, which can take days of processing, gives users the illusion that their queries run almost instantly. Unfortunately, there's no simple or automatic way to convert a live transactional database into an optimized CD-ROM database.

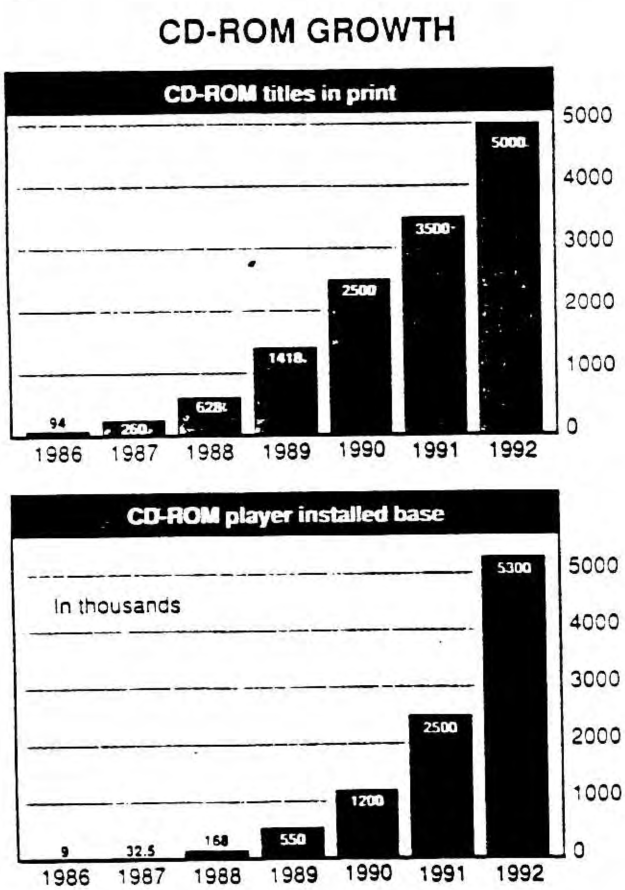


Figure 4: *The global rate of growth in both the installed base of CD titles in print (top) and players (bottom) since the inception of CD-ROM technology has been significant and steady. The numbers shown here, courtesy of InfoTech, represent both commercial and in-house units. Note that InfoTech's more conservative accounting treats data sets that may span many discs as single titles.*

Publishing for Posterity

CD-R clearly threatens WORM as an archival medium, although in the short run WORM's chief competitor will be MO (magneto-optical) technology (see the text box "Optical Flavors" on page 134). CD-R's compelling advantage is that you can read an archived disc in any CD-ROM player.

Creating that disc is complicated by three factors, however. First, you've got to translate the source file system (DOS, Mac, or Unix) into the neutral ISO 9660 file system that is the CD-ROM standard. Drives often provide formatters that convert on the fly as they write to the disc, but they may not handle some subtle naming conflicts. Hyphens are legal in DOS, for example, but not in ISO 9660.

Second, you must feed the drive a continuous stream of data. A dual-speed recorder, which fills a disc in a half-hour, needs a sustained 300 Kbps. That's only achievable from a fast hard disk in a controlled environment. Archiving a network drive must therefore be a two-stage process—one transfer to the CD-R workstation's hard disk, and then another to the CD-R drive. Finally, you'll want enough data on hand to justify the use of an expensive piece of media that can hold upward of 600 MB. Most CD-R drives don't yet support append operations (i.e., multisession recording), and none allow incremental file-oriented updates.

Other issues are the quality and longevity of CD-R media. NTIS's Betts found a distressing number of uncorrectable errors in early batches of CD-R discs he tested. Admittedly, Betts uses Enterprise Corporation of America's (West Des Moines, IA) CD-CATS (CD computer-aided testing system) to ferret out tiny defects that might well escape the notice of casual users. Still, "if Reed-Solomon [CD-ROM's error correction code] won't read," he says, "it is by definition not a CD-ROM." The CD-ROM specification requires functional error correction. Betts adds that the situation is improving rapidly.

Longevity is an open question for both CD-ROM and CD-R media. Since CD-R discs are burned, not pressed, they are more vulnerable to heat. Manufacturers also warn users to handle CD-R discs with greater care than CD-ROMs, implying that they're more fragile. But assuming that you take reasonable precautions with regard to heat and handling, there's no reason to think CD-R discs won't endure for 10 years, 25 years, or longer. It's even possible they'll outlast CD-ROMs, says Digipress (Louisville, KY) general manager Denis Oudard, an expert in archival CD-ROM. (His company's Century-Disc, a specialized $495 CD-ROM made of tempered glass, aims to last hundreds of years.) "The Achilles' heel of conventional CD-ROM is the aluminum reflective layer," says Oudard; it oxidizes too easily. CD-R discs use gold, which is more stable.

You don't have to own a CD-R drive to archive your data onto a CD-ROM. If it's something you'll need to do only once, you can

Optical Flavors

ANDY REINHARDT

CD-R offers a mix of drawbacks and advantages relative to its optical competitors WORM and MO (magneto-optical). With incremental, file-oriented write capability still on the drawing board for CD-R, it loses out to MO and WORM for interactive on-line use. It also loses out on price, with an average drive costing just under $10,000. WORM drives average $3800, while MO drives average roughly $4500. CD-R's great strength is standardization unmatched by any other form of removable media. Magnetic options such as Bernoulli and Syquest are proprietary, and the WORM market has been plagued by noninterchangeability. CD-R users can rest secure in the knowledge that archived data will be readable on any CD-ROM player.

In the short term, however, it's MO rather than CD-R that is usurping WORM's archival role. Researcher Freeman Associates' (Santa Barbara, CA) vice president Robert Abraham projects that sales of 5¼-inch WORM drives will plunge from 40,000 units this year to almost none in 1997, while 5¼-inch MO drives will grow from 174,000 to 400,000 units over the same period.

Competition for WORM will come from two kinds of "multifunction" MO drives. Hewlett-Packard, Sony, Hitachi, and Maxoptix sell drives that use standard 5¼-inch MO media but allow users to designate certain disks for write-once use only.

Critics charge that by merely emulating WORM, these drives don't offer the same degree of permanence and security as true WORM. Abraham says some people insist on using only ablative media that record permanently. Responding to that need, Pioneer and Laser Magnetic Storage sell drives that accept the same write-once media as their traditional WORM drives and also read and write 5¼-inch MO disks. Abraham says these drives are a good option for users who insist on true WORM but sometimes want the flexibility of rewritable media.

As CD-R drives and media get cheaper, recordable CD technology is likely to capture a growing share of the archival market. Early applications will have to be batch-oriented, however. "When you write to a CD," says Cris Simpson, an optical memory engineer with Pioneer Communications of America, "you have to lay down a big hunk of data because of the overhead associated with each session." File update capability for CD-R awaits completion, and then widespread acceptance, of the Frankfurt committee's ECMA (European Computer Manufacturers' Association) 168 specification.

Where CD-R is not expected to hurt WORM is in the specialized market for high-capacity storage. The only WORM drives Sony now sells are the 12-inch variety, which offer capacities of 6.5 GB and are often packaged in huge jukeboxes. "CD-R doesn't affect the regular WORM market for us," says Alan Sund, Sony's marketing manager for CD-ROM drives. "It's a totally different product and market, a whole different range of capacity."

Andy Reinhardt is BYTE's West Coast bureau chief. You can reach him on BIX as "areinhardt."

send a tape to your local CD-ROM copy shop and have it make a CD for you. Walnut Creek CD-ROM (Walnut Creek, CA), for example, accepts 8-mm and QIC (quarter-inch cartridge) tapes and charges $195 for overnight conversion to CD-ROM.

Client-Server CD-ROM

The push for a generic client-server model appropriate for CD-ROM has two motivations. The most critical need is to slow the proliferation of user interfaces that comes with a growing number of titles. My local college library has four CD-ROM stations, each with a different title. Because four different vendors produce those titles, there are four different user interfaces to master.

SFQL (Structured Full-Text Query Language), CD-RDx (CD-ROM Read-Only Data Exchange), and DXS (Data Exchange Standard) are among the proposed standards that promise just such consolidation. The viewing application will play the role of client, issuing requests to a data source that acts like a server. The benefit to users will be enormous. Unfortunately, only a few CD-ROM tool vendors have yet embraced any of the proposals now on the table. Many more, including Dataware and Silver-Platter, use a client-server model internally and are waiting for a clear winner to emerge.

A second reason for the client-server model is the need for effective remote access to CD-ROM stations. A file-based CD-ROM application usable on a 10-Mbps LAN grinds to a halt when you connect to those files at 2400 bps. A server-based CD-ROM, on the other hand, could communicate effectively with a remote client even over a slow link. I could dial into the CD-ROMs as I dial into my library's Ultrix-based on-line catalog.

None of these tantalizing future prospects should obscure the central message: Today's CD-ROM, warts and all, is often the medium of choice not only for commercial publishers but also for many forms of corporate communication. Thanks to CD-R, that choice just got a whole lot easier. The history of personal computing has shown over and over that when a powerful technology appears on the desktop, users seize it and proceed to change the world. Here we go again. ■

Editor's note: *BYTE news editors Patrick Waurzyniak and Ed Perratore, West Coast bureau chief Andy Reinhardt, senior news editors Gene Smarte and Tom Halfhill, and executive editor Rich Malloy also contributed to this article.*

Jon Udell is a BYTE senior technical editor at large. You can contact him on BIX as "judell" or on the Internet at judell@ bytepb.byte.com.

ADCP ISSUES

I. Hardware
- transducer and deck unit
- computer
- navigation inputs (gyro, GPS)
- 3DF GPS

II. Software
- profiler firmware
- data acquisition system
- user-exits (navigation, time)

III. User Support
- upgrades
- maintenance
- calibration and testing
- documentation

RVTech - standards and modes of operation

"de facto" standards for hardware and software

- RDI profiler
- RDI DAS
- Navsoft
- PC-based system

Modes of operation

- ADCP primary
- ADCP secondary
- routine underway collection

"fuzzy" standards: user support

- software and firmware upgrades, documentation of version numbers and updates
- serial number of transducer, deck unit
- history of transducer inspection
- history of deck unit refurbishment
- installation:
  - acoustic window
  - heading offset
  - calibration information
  - data quality information
- gyrocompass calibration
- gyrocompass compensation procedure
- navigation inputs
- time source

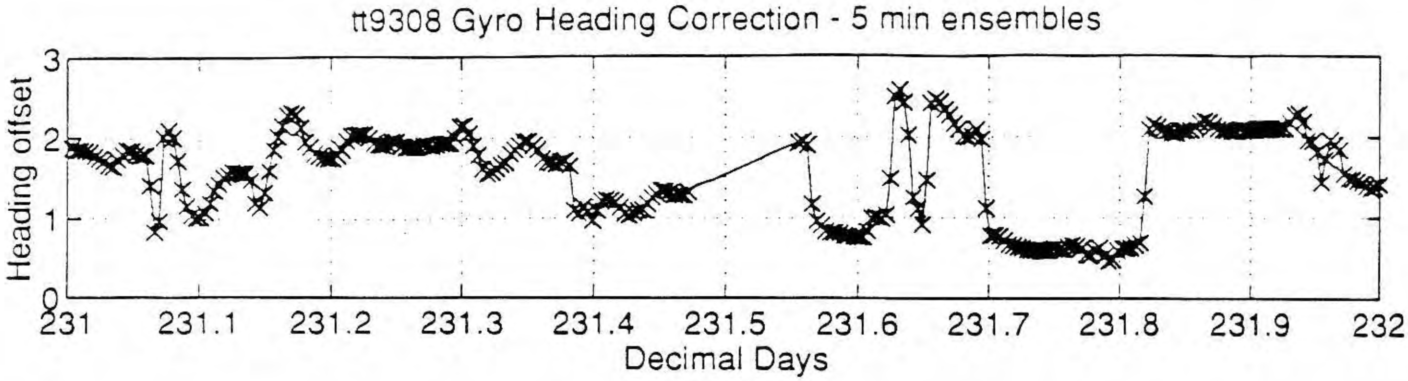
Installation

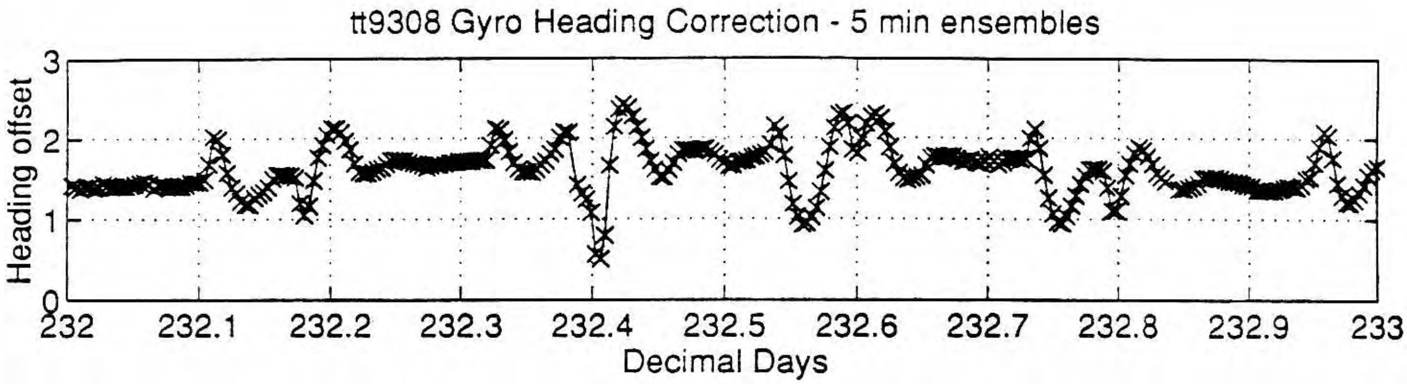
- low noise region of hull
    - minimize bubbles
    - away from propellers
  - window
    - faired pod

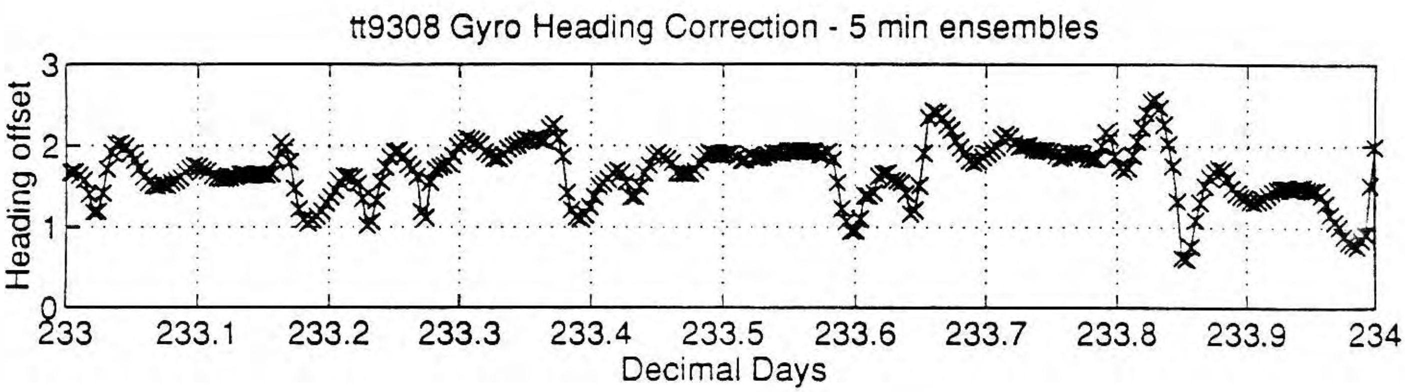
- ease of installation

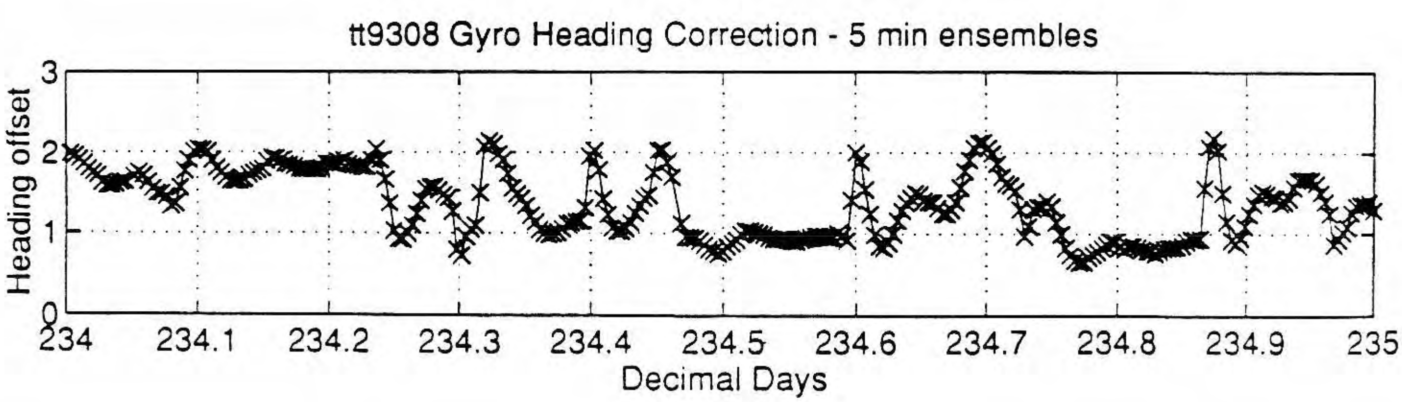
- ease of inspection

- testing of installation

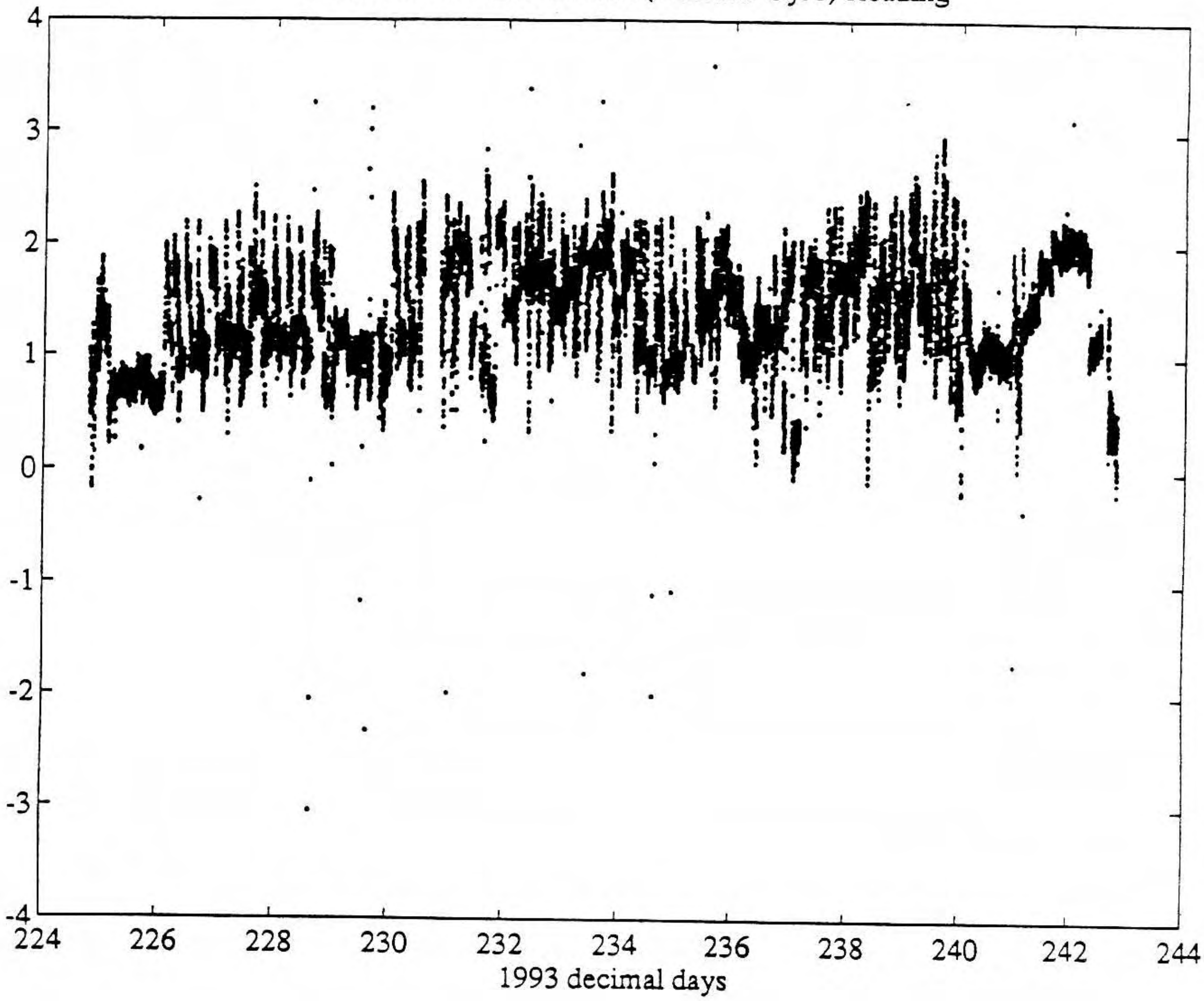

tt9308 Gyro Heading Correction - 5 min ensembles
Heading offset
0
1
2
3
231
231.1
231.2
231.3
231.4
231.5
231.6
231.7
231.8
231.9
232
Decimal Days

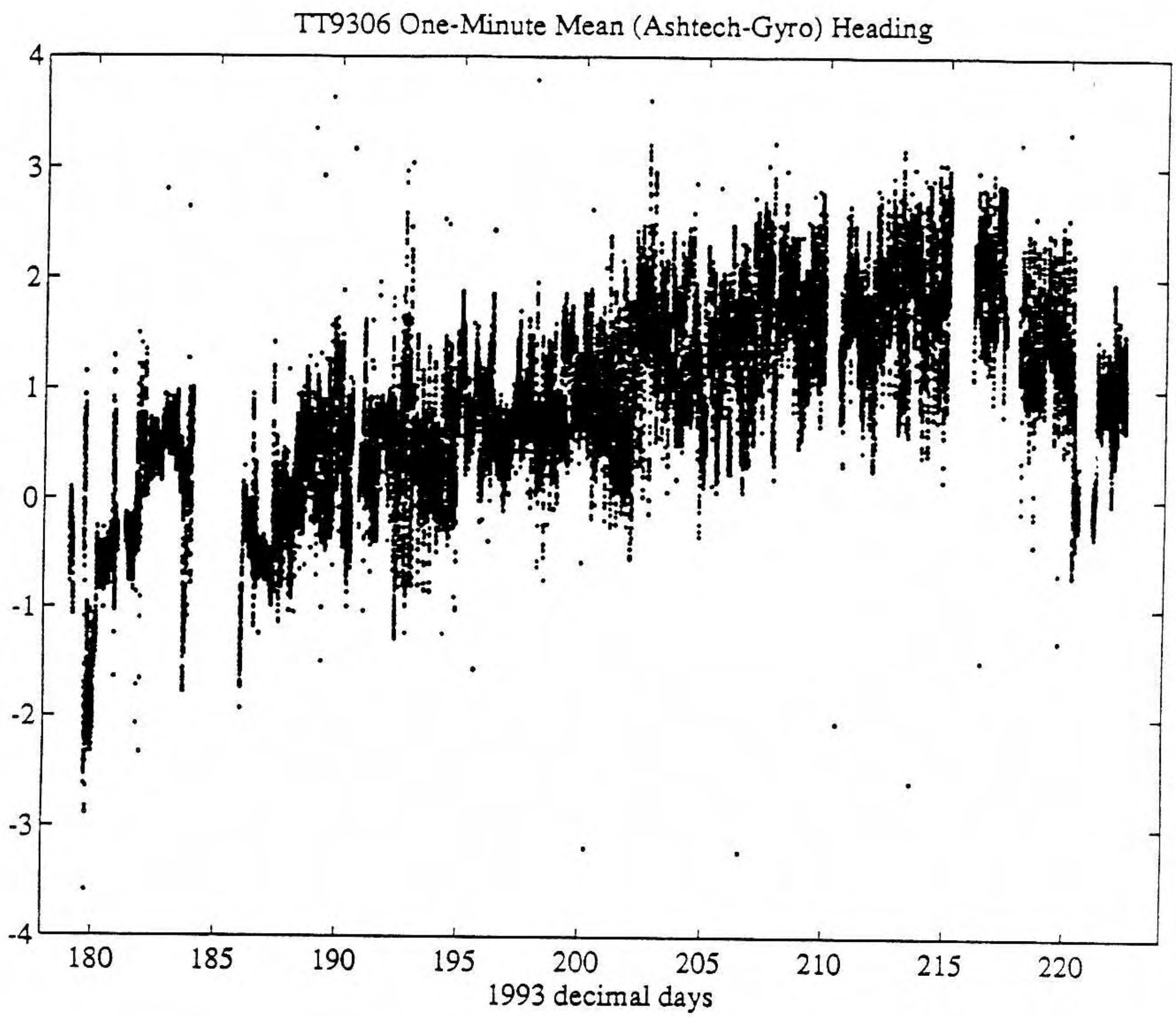

tt9308 Gyro Heading Correction - 5 min ensembles
Heading offset
0
1
2
3
232
232.1
232.2
232.3
232.4
232.5
232.6
232.7
232.8
232.9
233
Decimal Days


tt9308 Gyro Heading Correction - 5 min ensembles
Heading offset
0
1
2
3
233
233.1
233.2
233.3
233.4
233.5
233.6
233.7
233.8
233.9
234
Decimal Days


tt9308 Gyro Heading Correction - 5 min ensembles
Heading offset
0
1
2
3
234
234.1
234.2
234.3
234.4
234.5
234.6
234.7
234.8
234.9
235
Decimal Days

TT9308 One-Minute Mean (Ashtech-Gyro) Heading

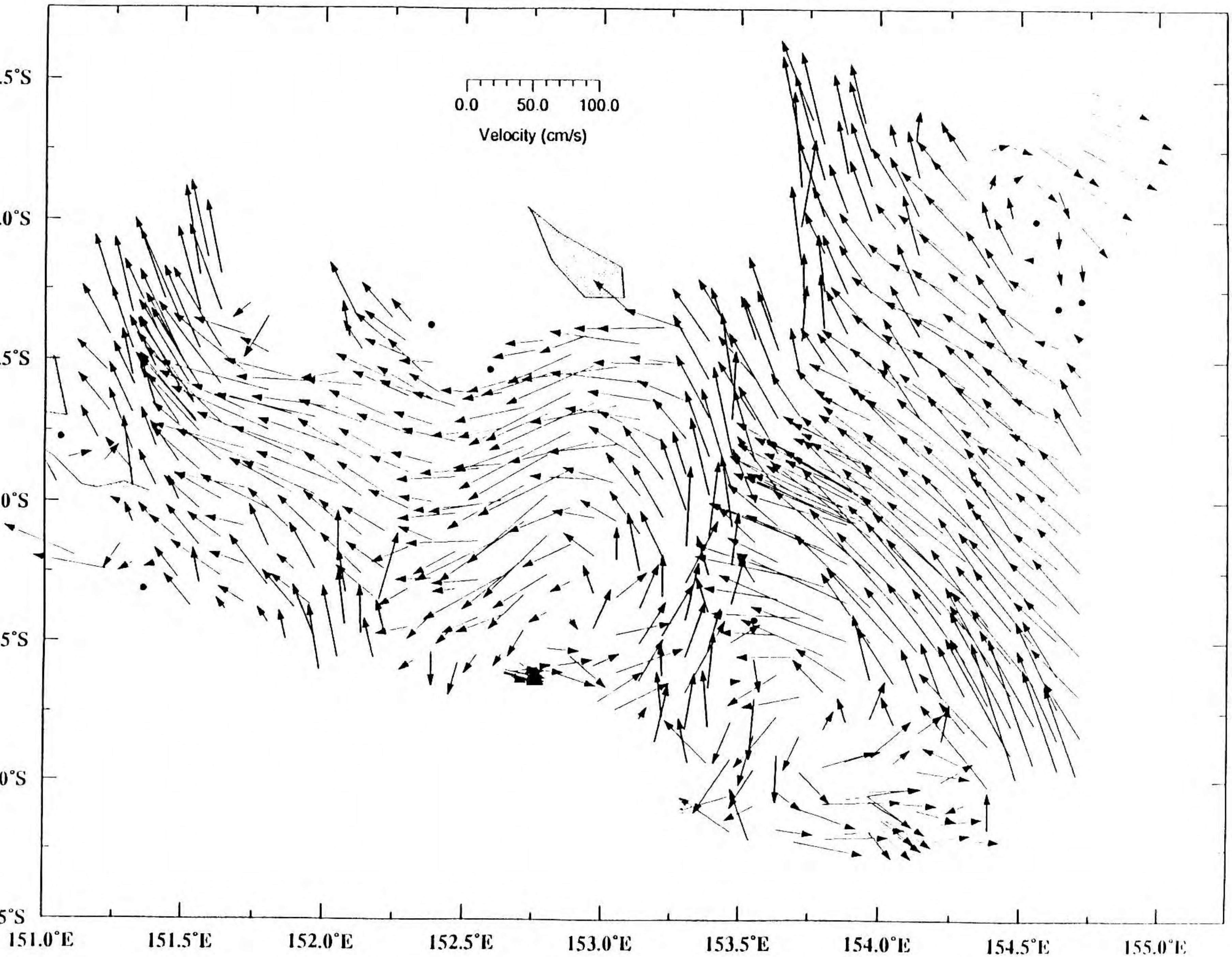




TT9306 One-Minute Mean (Ashtech-Gyro) Heading
4
3
2
1
0
-1
-2
-3
-4
180
185
190
195
200
205
210
215
220
1993 decimal days

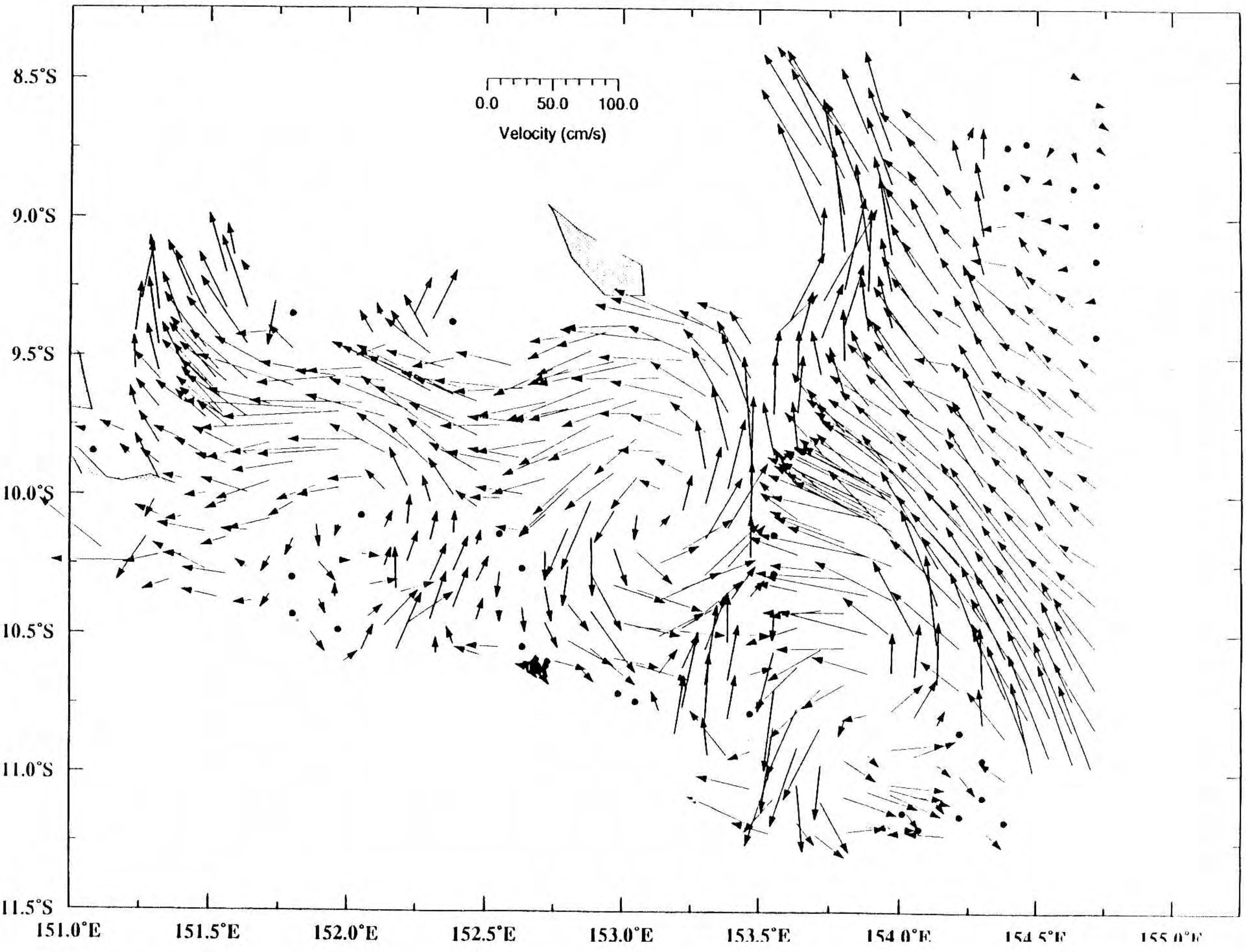
MW9304: 93/04/25 to 93/05/21

Layer: 25m to 75m



MW9304: 93/04/25 to 93/05/21

Layer: 125m to 175m



SeaNet Shipboard LAN Guide

This guide is prepared as part of the SeaNet program plan.

Prepared by:
Rex A Buddenberg
budden@shell.portal.com
for Joint Oceanographic Institutions
August 1993

Introduction

This Guide outlines the Local Area Networking that oceanographic ships need in order to effectively participate in SeaNet. Here we'll discuss the shipboard parts of SeaNet -- routers, network management and the LAN within the ship. Most of the guidance here is independent of the decisions made in the radio WAN portion of SeaNet. The reasons we're putting this early on the agenda:

--> there is no development required in LAN technology on order to meet our SeaNet needs. Everything can be purchased off the commercial shelf today.

--> ship schedules dictate installation schedules. A LAN that conforms to the larger vision requires an availability period to install. Institutions may choose to work ahead and put the installation job into a planned availability period.

--> real value can be realized from an organized shipboard LAN, even without the radio WAN interconnect.

--> only a small portion of the shipboard LAN will be funded under SeaNet. Contributions may come from several sources; this Guide should help each institution make each LAN component contribution add to the larger structure within the ship.

--> the Guide will help those people planning a pre-packaged experiment by defining an interface that the prepackaged experiment can 'plug into' much like the experiment must 'plug into' ship's electrical power.

Order of business

This Guide presents two LAN outlines:

--> The first is a 'visionary' outline -- it is expansive and intended to meet a large variety of intra-ship communications requirements. These include integrating the vessel's command systems (such as navigation and ship control) as well as the purely scientific ones. This outline contains a great deal of growth capability and should meet the ship's needs for some time to come.

--> The second outline is a 'minimalist' one. This is the minimum necessary for a ship to participate in SeaNet without the LAN becoming a weak link in the system. The minimalist outline should be installable without requiring an availability, and can likely be done with ship's force personnel.

The actual installation in your ship will no doubt fall somewhere in between. Detailed installation standardization is not the objective -- rather we're after interoperability.with a dollop of commonality.

The Visionary LAN outline.

There are several reasons that support this version of the shipboard LAN:

--> geological and geophysical oceanography deal with volumes of data that justify this kind of expansive view.

--> the ship is equipped to deal with the data requirements of most any experiment likely to be brought aboard.

--> the LAN can be thought of as part of the ship -- a utility -- rather than something carried by the ship.

--> The local area networking substrate can support command functions, such as navigation and ship control, in addition to the scientific data communications support. (There are some commercially available ship control and navigation packages built on high availability local area networks).

General layout.

Think of the shipboard LAN as a backbone with several stubs. The backbone is a dual-ring Fiber Data Distributed Interface (FDDI) LAN with a cable plant that visits all spaces in the ship likely to host ship control functions or science functions (effectively just about everywhere). In practice, the backbone cable plant should be the permanent installation and therefore should be the only LAN cabling that penetrates watertight bulkheads.

At several places around the FDDI ring, insert a hub to provide easy means for interconnect to end systems. The hubs can support a direct interconnection (e.g. a sensor plugged in directly) or other LAN segments (e.g. an ethernet segment that, in turn, has end systems attached to it).

This arrangement allows further expansion by adding more LAN segments connected either to the router or to one of the hubs. These extensions could include wireless LANs and subsurface (acoustic) LANs.

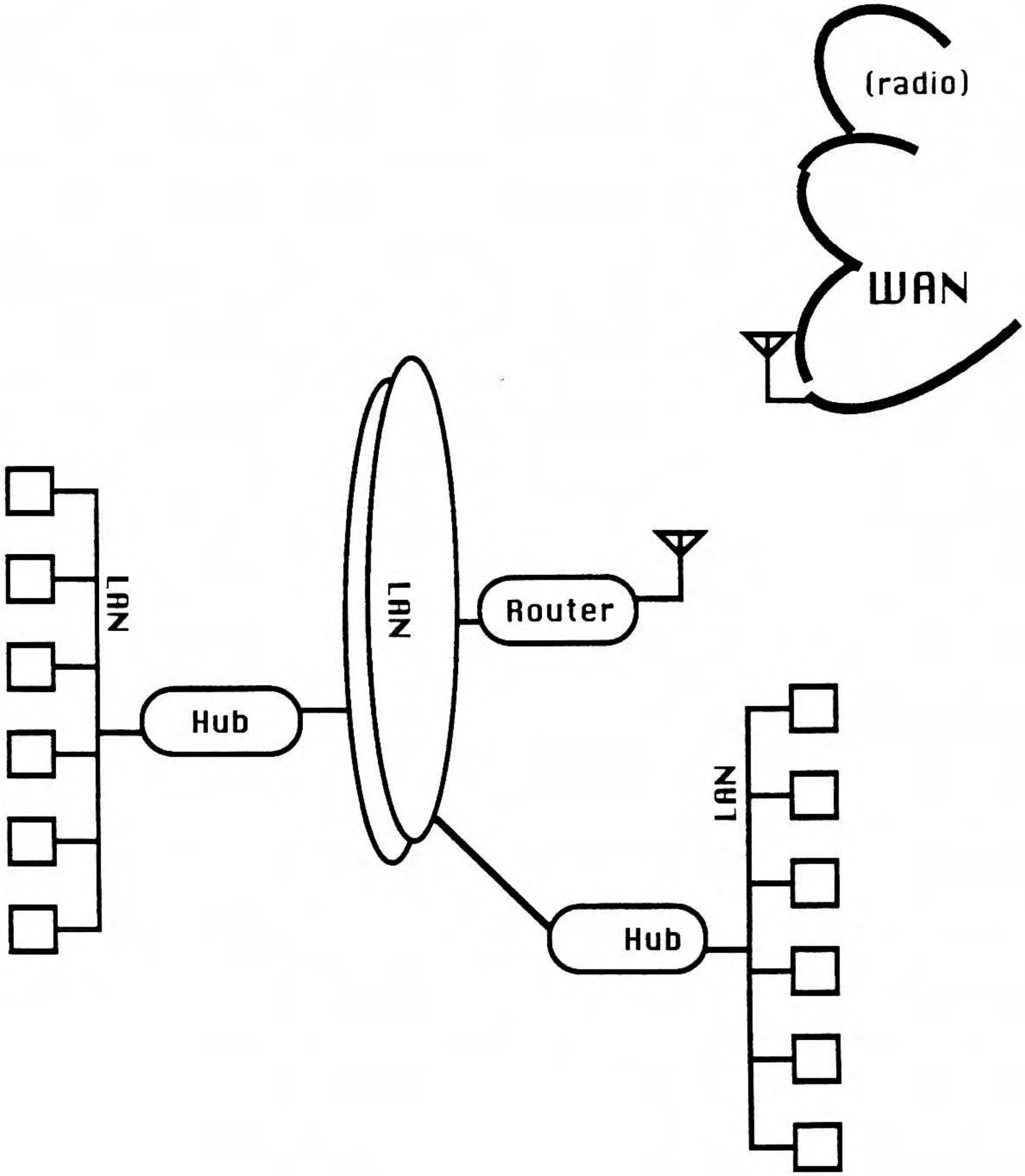


Figure 1 Shipboard LAN vision

Backbone segment -- SAFENET

We recommend that the backbone FDDI LAN be patterned after the Navy's SAFENET standard to eliminate single points of failure that cause downtime. The rest of this section describe this LAN configuration and tailoring[1]. SAFENET uses the Fiber Data Distributed Interface (ANSI X3T9.5, FDDI) which provides several attributes that we need:

--> ability to provide two, counter-rotating, rings. This meets our initial high availability principle of redundant connectivity

--> ability for each node to detect downstream failures. When this happens, the node 'wraps' the data back onto the second ring. This isolates a failure allowing the network to continue functioning. Because the data is present on both rings, the surviving nodes can use either connection. This makes the crossover action reliable, meeting another high availability principle.

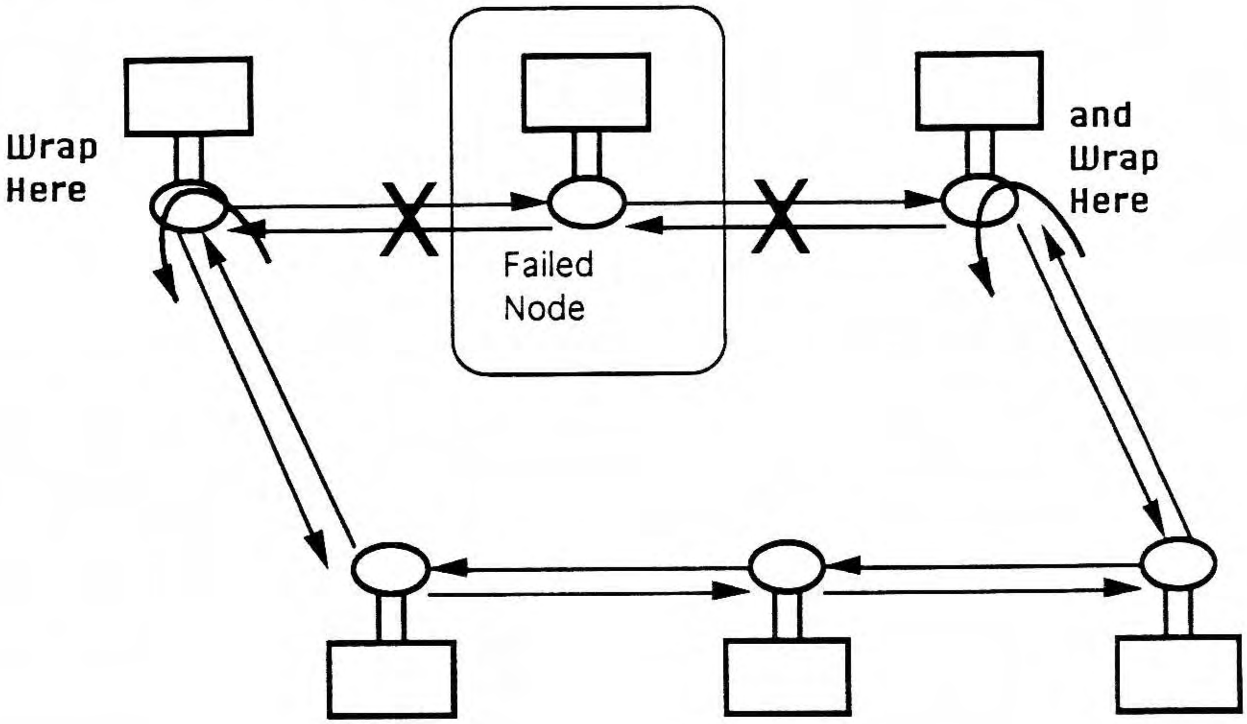


Figure 3 FDDI Ring Wrap

--> the FDDI standards include a strong station management protocol allowing a network manager to detect failures and direct maintenance activity effectively.

---

[1]SAFENET (Mil Std 2204 and Mil Hdbk 818-1) started life as an acronym: Survivable Adaptable Fiber optic Embedded NETwork.

The following recommendations are drawn from SAFENET. They are not part of the base FDDI standard -- they qualify as tailorings that you must request in order to meet your high availability requirements. Caveat: some of these will sound like overkill -- SAFENET is designed to support combatant ships and their mission critical systems. Tailor down to your needs.

--> dual rings and dual attachment stations. The standard supports dual rings. Make sure you get both. In FDDI lingo, specify 'dual attachment stations'. A single attachment station represents a single point of failure where a fault in it will take the network station out of communication with the rest of the network.

--> concentrators. With the possible exception of your routers, all hubs and other equipment should be attached to the LAN either through concentrators or optical bypass switches. This insulates the LAN trunks from the effects of equipment failure

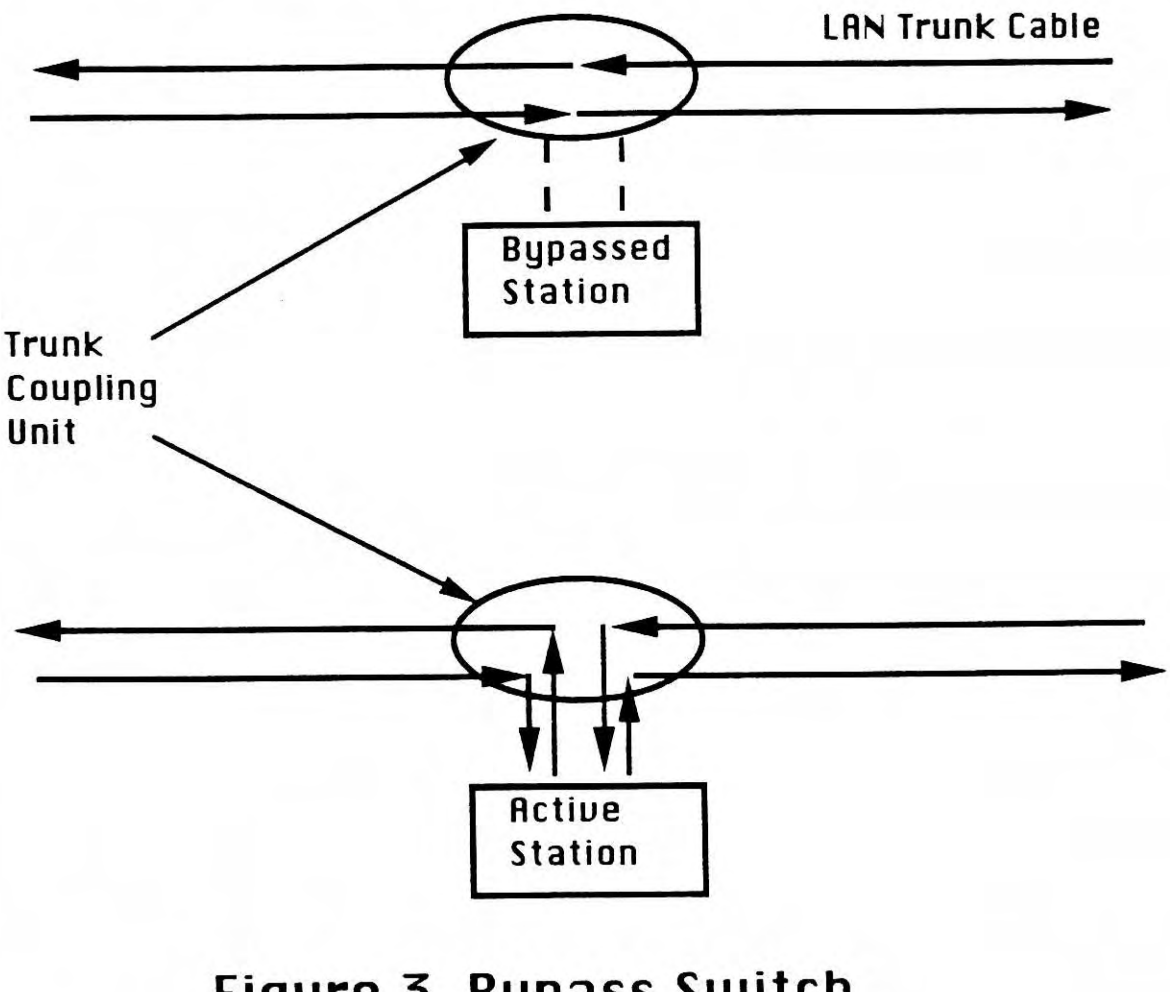


Figure 3 Bypass Switch

--> trunk coupling units with optical bypass switches. These switches either pass data around the station when the station is disabled or allow the data to pass to the station when it is in operation. Bypass switches allow end systems to be powered off without bringing the network down.

--> dual Media Access Controllers (MAC). This requires a doubling of hardware in each connection node but eliminates the node as a single point of failure -- if one MAC fails, the other can function until repairs are made. Further, dual MACs allow easier testing of new subnetworks before insertion into the main network which will be important as your network grows. Some of the board manufactureres place a MAC on each of a pair of boards and support 'hot swap' so you can replace a defective MAC while the system is up and running.

--> splices. Normal commercial cable and installation practice for FDDI is generally acceptable. FDDI cable can be run through stuffing tubes just like electrical power and copper communications wiring. In general, fusion splices are more durable and entail less loss than connectors, but require an experienced installer and specialized equipment.-- you should generally use them for the FDDI trunk cabling (from one trunk coupling unit to the next).

--> connectors. Connectors are suitable for local connections -- between trunk coupling units and concentrators or hubs -- especially where equipment is likely to get relocated. 'ST' (bayonet[2]) or 'MIC' (plug-in) connectors are used for FDDI LANs and either is appropriate (screw-in connectors are prone to vibrate loose). You should, however, be consistent and settle on one or the other. Adding connectors to fiber optic cable requires similar expertise and equipment as splicing but you can finesse this by buying pre-connectorized cables.

--> LAN cable topology. There is no benefit to diverse routing of the FDDI trunks because of the characteristics of ring 'wrap' when a fault occurs. Therefore, simply install cable in the ship that contains at least two fibers, one for each ring. In naval (or ex-naval) ships, one place to start is to consider the cable run for the degaussing coil around the ship.

Extra pairs of fibers in the cable for expansion is a good idea -- the cost of the media is insignificant compared to the installation labor. And the extra pairs allow you the option of segregating subnetworks by separate attachments to the router, for the sake of the network manager's sanity.

---

[2] 'ST' stands for 'snap/twist'.

--> FDDI cable material. Mil-C-85045 and Mil-C-24643 (referenced in SAFENET) contain general specifications for shipboard fiber optic cable. While the MilSpec cable is optically identical to commercial fiber optic cable, the spec contains several low-smoke and chemical makeup provisions that affect shipboard safety, particularly in the event of fire.

Additional FDDI features

There are several features that fiber optic connectivity brings to your shipboard LAN that copper cannot. Therefore, we recommend fiber rather than one of the derivative copper-based implementations of FDDI:

--> Fiber optic cable does not emit electromagnetic interference nor is it affected by it. This is a significant consideration when you have lots of other electronics, especially radios, around[3].

--> Shipboard fiber cable and most fiber cable for buildings is completely dielectric[4]. This means that nothing in the cable can transmit electricity -- no ground loops.

--> And fiber offers greater growth potential -- the current FDDI standard has a capacity of 100M bits per second but developing technology will offer higher rates over the same cable plant[5]. It's less likely that a copper cable plant will be able to handle greatly increased capacities. Given the difficulty of installing any kind of permanent cable plant in a ship, it makes all kinds of sense to install a cable plant that is unlikely to be outgrown over the expected life of the ship.

---

[3]This characteristic also makes fiber much more difficult to wiretap than copper.

[4]Some versions for exposed locations and direct burial have metal sheaths.

[5]Several developing LAN technologies can use the same cabling plant. Among these are:

--> FDDI-2. Enhancements of the current FDDI standard.

--> ATM, Asynchronous Transfer Mode. Likely the next generation WAN standard to overtake X.25, Frame Relay and SMDS. Some (perhaps overenthusiastic) people claim that ATM will also overtake existing LANs and provide connectivity to the desktop.

--> FiberChannel. A high speed LAN effort for connecting supercomputers.

Stub LAN segments -- Recommendation

Recommend that stub LAN segments use ethernet, 10Base-T (aka unshielded twisted pair or UTP) cabling. It's popular, inexpensive, easily reconfigured using ordinary tools, and will do the interconnect job.

Ships may also want to stock ethernet, 10Base-2 (aka thinnet) adapters and 50 ohm coaxial cable. This is mostly for installed base reasons. With inexpensive adapters, both standards can be accommodated and rather freely mixed.

There are several feasible alternatives discussed in the Rationale appendix but we recommend avoiding a wide proliferation of LAN types -- not for interoperability reasons but for logistics and network management reasons.

With the installation of a few hubs around the ship, it becomes easy for individual sensors and computers or entire prepackaged experiments to interconnect to SeaNet just like R2D2.

Devices on the LAN

Beyond the cable plant, several devices should connect to the LAN to provide necessary services to users. Below are some requirements to consider when selecting products.

Internetworking -- the routers

A router is a network interface device that has LAN and WAN connections to it including the obviously necessary radio WAN portion of SeaNet. It is capable of switching data from one LAN to another LAN or to a WAN, depending on the address attached to each chunk or packet of data. This is the 'glue' that holds our different kinds of networks together.

The router should be installed in the ship's communication spaces where it can be connected to the radio-WAN equipment.

Interim spec. For a variety of reasons, SeaNet should buy routers from a single source for provision to all SeaNet ships (particularly in the first generation of the SeaNet program). This makes selection of a router a task for the SeaNet-implementing organization rather than the individual institution. But an institution (or ship) may not be able to wait, so here is some interim guidance:

Use the following guidelines when buying and installing your routers[6]:

--> the router should, at a minimum, be able to interconnect your shipboard LAN and your radio-WAN service(s). It should also be able to interconnect any administrative LANs (usually either IEEE 802.3 ethernet as discussed above).

--> the routers must be able to handle the Network Layer protocols selected. Virtually all routers handle Internet Protocol (IP)

--> There are several protocols that routers may use to pass the routing tables amongst themselves; you need one of them. Open Shortest Path First (OSPF) is the current generation, supplanting the more common Routing Information Protocol (RIP). Get OSPF if you have no installed base worries.

--> the router should be a 'managed device' conforming to the network management protocol. Specifically Simple Network Management Protocol, Version 2 (SNMPv2). Routers are probably the single most important components to be managed in order to meet our high availability requirements. In SeaNet it will be vital that on-board network managers be able to 'view' the network through the management capability. Further, we need to be able to provide an expert 'look over the shoulder' from network managers ashore.

--> the router should be equipped with an Uninterruptable Power Supply (UPS). Your mileage may differ regarding ship's power philosophy and reliabiliy. Author's recommendation is a UPS capable of providing an hour's lights-out operation.

--> capacity -- a non-requirement. SeaNet capacity will almost always be paced by the capacity of the radio-WAN. Even the lowest capacity routers on the market can work faster. The packets/second specifications for routers are interesting reading, but not important to the SeaNet requirement.

Network Management Station

The SeaNet program intends to supply a network management capability as part of the ship-station package. Like the router recommendation, this is interim guidance for those who cannot wait.

Each ship needs the capability to manage the LAN within. Install network management software on one of the workstations

[6] Bridges, which are good for interconnect multiple LANs, are not appropriate for this job of interconnecting both LANs and WANs.

that is accessible to the human network manager, whoever that person(s) may be.

The network management software, and probably the host hardware that runs it, should be centrally procured. Like the router, there are many logistic reasons why these should all be alike.

Network management software should conform to the Simple Network Management Protocols (SNMP).

As a rule of thumb, install the network management station near the router. There is no reason why the network management software cannot be installed on more than one workstation if different people need a view of the network.

The network management station should be UPS supported.

Hubs

Commercially available hubs range from the 'harmonica' with a handful of plugs to large and 'stackable' hubs designed to be installed in office building wiring closets. We're envisioning something at the more modest end of the scale that:

--> interconnects to the FDDI backbone,

--> provides a half dozen ethernet ports (which typically have one to five end systems attached to each),

--> possibly some serial interfaces (see communication controllers below),

--> are managed devices (SNMPv2, see the network management section).

--> recommend rack mounted devices,

--> at least half of the hubs installed in a ship should be supported with a UPS. With the optical bypass switches installed in the backbone, this should be enough to keep the FDDI backbone operating when ship's power fails.

Communications controller or terminal server

Install a communications controller where you need to interface oceanographic sensors, radionavigation receivers or other 'non-network savvy' devices to the ethernet LAN. In commercial practice ashore, terminal servers are used to support dial-in modems, dumb terminal connections and shared printer connections.

- server must include TCP/IP TelNet support -- including reverse connection.

- DB-25 ports for RS-232 connection (this should meet 90% of any requirements)
- SNMP management, same as the router and hubs.
- 10BaseT network connector to connect to the stub LANs

Note that the communications controller function can be achieved by adding I/O boards to a workstation. You may also find serial I/O capabilities on hubs (in addition to the network ports) that allow them to meet your needs. Your approach may be a matter of economics and taste.

The minimal shipboard LAN outline.

There may be several reasons why an elaborate LAN like the visionary outline above may be unnecessary or impractical:

--> inadequate funding,

--> lack of an availability period for installation,

--> the vessel doesn't have the remaining life to justify,

--> the type of oceanographic work and projects carried aboard does not justify.

This section outlines the minimum necessary for a ship to participate in SeaNet.

Again for logistic reasons, SeaNet intends to procure the ship station (including router and network management station) minimum centrally. Brand name uniformity for routers and network management stations is almost mandatory for logistics and (human) network manager sanity reasons.

You need:

--> the router, as described above.

--> enough LAN connectivity to reach the end systems you need to attach to SeaNet. Recommend 10Base-T (ethernet over unshielded twisted pair).

--> at least one workstation capable of supporting the protocols, network management software, and electronic mail user agents.

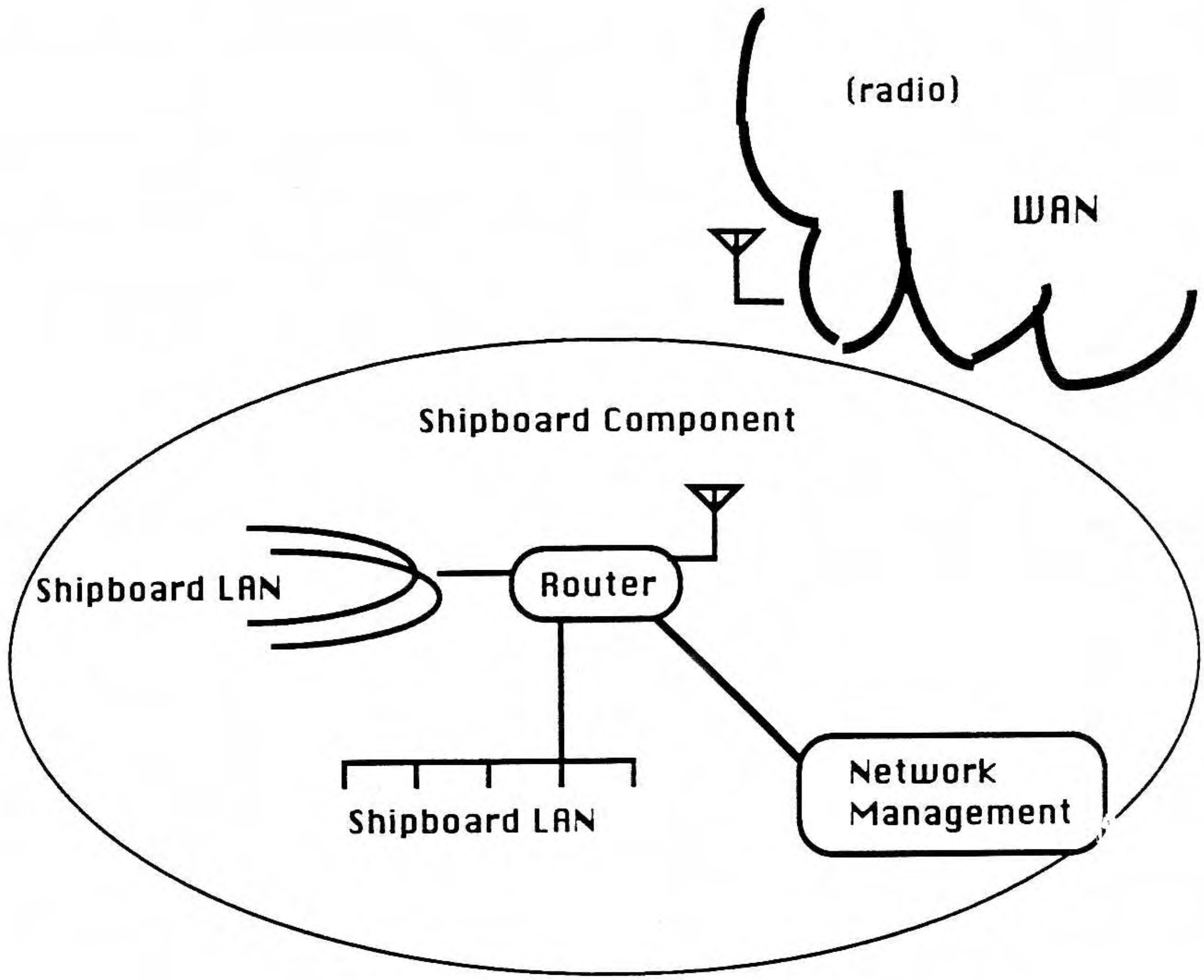


Figure 4 Routing and Network Management

(insert picture)

This is the minimum that preserves the modularity for growth, and for centralized development of the radio-WAN (the part pertinent to the shipboard LAN being the router). Recommend you avoid installing the cable plant permanently in the ship and avoid penetrating watertight bulkheads.

We feel strongly about the network management and UPSs and urge their inclusion. If you omit them, understand the risks you run.

Appendix A -- Network management

Network management includes the following:

--> network fault notification/isolation. For our SeaNet, this is the most important part as it targets the maintenance activity critical to keeping the system operating.

--> managing security Because SeaNet inherently uses shared media, and a shared radio media at that, it's important to keep some control to exclude the riffraff and prevent (either malicious or accidental) crashing of the network through intrusion.

--> understanding and configuring network performance. Who are the heavy users of the network? Since the community network will have a lot of varying usage, it may occasionally be necessary to restrict access to critical functions. An understanding of the way your network operates is key to targeting expansions of capability and managing the networks growth over time.

For SeaNet, fault detection, isolation and repair is by far the most critical problem -- keeping the network in a functioning state. A couple tips on router installation:

--> routers should routinely be provided with uninterruptable power supplies.

--> the network management stations will work best if they are located both physically and logically close to the routers.

How do network management systems work?

Background. This appendix is for education. Few textbooks, and little of the advertising hype, explain how the network management protocols work.

In its construction, Simple Network Management Protocol (SNMP) is similar to the other Application Layer services such as electronic mail and file transfer. And the players are typical client/server:

--> server. The network manager is a software product that resides in a workstation operated by the human network manager. As a rule of thumb, locate this is in a communications space, near your router(s).

--> clients. The managed devices are conventional network devices like routers and computers that have 'management agents' in them. These management agents are the chunks of software that are able to respond to the server in the language that they both understand. Note that the class of clients is open ended -- other communications devices like modems, radios, ... can be added. And non-network devices -- end systems -- like radionavigation receivers and CTDs can also be brought into the management regime.

The 'language' of SNMP consists of three message types:

1. Get. The network manager queries a device. The device returns a value -- a GetResponse -- to the manager. Gets allow a manager to monitor device activity -- traffic reports, configuration reports, temperature, ...

2. Set. The network manager orders a change in the remote device. This allows control and reconfiguring, not just monitoring.

3. Trap. While gets and sets are initiated by the manager, traps are initiated by the remote device. Traps are alarm conditions (power fail, unauthorized access, hardware failure, traffic volume exceeding a predetermined threshold, etc)

The manager and management agents are routinely not manufactured by the same company. So the data dictionary must be agreed upon as an open standard. In SNMP, an object is the formal definition of a single element of management information -- a data atom. This collection of objects is known as the Management Information Base (MIB). In addition to the (currently ~200) standard objects, experimental MIBs, evaluation and draft MIBs and private MIBs are all allowed and provided for. Note that there is a standardization system and a repository.

You should make your supporting vendors responsible for delivering equipment with management agents in them, and in creating and registering the interoperable MIB information. Require demonstration of interoperability with your management server in your acquisition process. Managed devices will cost you more in initial procurement outlay than unmanaged devices. But since management is central to controlling our fault tolerance problem, the extra cost is worth it. And you're likely to get the expense back in repair technician economies.

Management security. SNMP happened quite fast in the networking world. The first edition of the standard lacked support for assuring authenticity of the gets, sets and traps described above. Consequently, a malicious intruder, if he can penetrate your network, can impersonate the network manager and wreak havoc with your network configuration[7]. The second generation of the SNMP

---

[7]This was recognized when the original SNMP standards were drafted, but the security problem took longer to solve in a multi-vendor interoperable standard. Meanwhile networks needed SOME management capability, even

standard, called SNMPv2[8], contains security provisions to prevent malicious reconfigurations The security scheme uses the same public key encryption processes described for electronic mail security.

Remote management. An ability to remotely manage networks and devices attached to it as a vitally important part of SeaNet. Good (human) network managers are hard to find; good network managers that will also go to sea are likely to be somewhat rarer. Further, SeaNet intends to extend to unmanned platforms (buoys, UAVs) as well. All of which justifies attention to the network management topic, including remote management.

Remote management can allow expert network managers ashore assist the afloat technicians in diagnosing and repairing problems and coaxing a balky system into operation. The SNMP standard includes a section on remote monitoring (called RMON in the standard) that permits this. Good idea for hubs and terminal servers Two features in particular:

--> remote probing is supported by the standard. A probe is a monitoring device that is attached to a segment of the LAN. It listens to the traffic on that segment and passively derives management information from that traffic. This information is bundled up and sent to a remote network manager which, in the case of SeaNet may be ashore.

--> the remote monitoring standards account for manager-to-manager interaction as well. Here a network manager ashore can collect management data from the network managers in several ships or drifters to provide an integrated picture.

From a LAN Guide point of view, this is interesting information that justifies the recommendations we've made above.

Appendix B -- Continuity of electrical power

Most shipboard personnel are familiar with the blackness and eerie silence that accompanies a shipboard power failure. Some of your colleagues ashore may not be as attuned to the problem. So, some more background information.

The computers, routers, hubs, ... on the shipboard LAN must be supplied with electrical power continuously.

First, everyday electrical power is often polluted with power

---

without the security.

[8]RFCs 1441-1452 contain the SNMPv2 standard.

sags, surges, spikes, noise, and in ships, cycle swings, that can cause computing errors and hardware damage. Shipboard power is often worse than shore power since there is no continental power grid to absorb wierd things that happen..

What we need is both a filtering capability to control the noise and a stopgap to keep everything running when the ship's hotel power fails or hiccups. This is the role of the Uninterruptable Power Supply (UPS). There are a couple acceptable alternatives.

The routers and hubs in the shipboard LAN portion of SeaNet should have a UPS. Here are some points to consider when procuring and installing:

--> the UPS should include power line conditioning features. These filtering requirements could be met with separate filters but since you need UPSs anyway, include the filtering features in the UPS.

--> UPS capacity should be adequate to carry the supported routers and hubs for about an hour. If you can't get ship's power back in that time, you've probably got worse problems than a network crash on your hands.

--> make sure you can assure yourself that the UPS is working correctly. When the outage occurs is not a good time to find out your UPS has tired batteries.

--> some UPSs have SNMP management agents in them so they can be included in the network management regime. Highly recommended for SeaNet -- especially on the router.

--> UPSs, like all equipment, fail too. So don't connect both prime and backup equipment to the same UPS and let it become a single point of failure.

Appendix C -- Stub LAN segments

Educational background on LAN standards. There are four standards for LANs that are formally recognized in the standards world. In addition to FDDI:

--> The most common is the IEEE 802.3 (known as Ethernet). 802.3 networks cannot be made fault tolerant[9]. Further, they cannot provide deterministic service which is a more subtle, but significant requirement in operational situations.

---

[9] Some vendors are offering fault tolerant systems using dual Network Interface Cards based on ethernet. But such solutions are not within the standards and are not supported by the network management regime.

--> The IEEE 802.4 (token bus) network is obsolete and quite rare now and suffers the same fault tolerance shortcomings.

--> The IEEE 802.5 (token ring) LAN has the potential for fault tolerance and deterministic delivery (the Navy started here with SAFENET) but little industry support for high availability applications. Because the 802.5 standard was supported by IBM, it tends to be most popular where some kind of IBM backfit is needed (a problem that doesn't seem to exist in the oceanographic fleet).

These three standards should not be used for your backbone LAN, but supporting any or all of them locally from the hubs is quite acceptable. Ethernet, of some flavor or another, appears to be by far the most popular. Either 10Base-2 (thinnet) or 10Base-T (twisted pair) are easy to work with and popularly supported. (There is little reason for obsolescent 10Base-5 (thicknet) or FOIRL (fiber) -- the attributes they offer should me met by use of FDDI, see the backbone LAN segment discussion above). For the sake of the sanity of the network manager, settle on one, probably Ethernet 10Base-T.

There are two next-generation Ethernet standards in process at this writing. Both purport to provide 100M bits per second.

--> one simply combines the existing Ethernet MAC (controllers have routinely been implemented in silicon for several years) with the FDDI physical layer standards. This is appealing for vendors with large installed bases as they can transparently implement the new standard on new workstations and it will still work with older workstations (at the old 10M rate) on the same LAN segment. The stochastic delivery characteristics of Ethernet remain.

--> the second proposal alters the MAC layer too in order to provide isochronous delivery characteristics. This is a 'bandwidth reservation' system that would allow things like full-motion audio or video delivery over the LAN. These performance characteristic improvements are gained at the sacrifice of installed-base compatability.

Niether of the proposed Ethernet standards address the fault tolerance characteristics that led us to recommend FDDI above.